

NBER WORKING PAPER SERIES

BIG LOANS TO SMALL BUSINESSES:
PREDICTING WINNERS AND LOSERS IN AN
ENTREPRENEURIAL LENDING EXPERIMENT

Gharad T. Bryan
Dean Karlan
Adam Osman

Working Paper 29311
<http://www.nber.org/papers/w29311>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021, Revised May 2022

Approval from the Innovations for Poverty Action Human Subjects Committee (IRB#0004745). This project was registered on the AEA RCT Registry (AEARCTR-0004617). For funding we thank DFID's PEDL, the Innovation Growth Lab and the Argidius Foundation. For project management and field support we thank Abdelrahman Nagy, Mohamed Omar, Mohamed Yousef, Sarah Wahby, Sara Serrag, Nader Kabbani, and Tarek Coury. We thank the management at ABA including Magdy Mousa, Motaz El Tabaa, Tarek Abdelaziz, Ibrahim Melouk, as well as support from Mona Zulficar and Hasan Ibrahim. Eduardo Medina provided superb research assistance. We thank Muhammad Meki and Simon Quinn for helpful comments. The authors retained full intellectual freedom to report and interpret the results throughout the study. All errors and opinions are those of the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Gharad T. Bryan, Dean Karlan, and Adam Osman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Big Loans to Small Businesses: Predicting Winners and Losers in an Entrepreneurial Lending Experiment

Gharad T. Bryan, Dean Karlan, and Adam Osman

NBER Working Paper No. 29311

September 2021, Revised May 2022

JEL No. D22,D24,L26,M21,O12,O16

ABSTRACT

We experimentally study the impact of substantially larger enterprise loans in Egypt. Larger loans generate small average impacts, but machine learning using psychometric data reveals that "top-performers" (those with the highest predicted treatment effects) substantially increase profits, while profits drop for poor-performers. The large differences imply that lender credit allocation decisions matter for aggregate income, yet we find that existing practice leads to substantial misallocation. We argue that some entrepreneurs are over-optimistic and squander the opportunities presented by larger loans by taking on too much risk, and show the promise of allocations based on entrepreneurial type relative to firm characteristics.

Gharad T. Bryan
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
g.t.bryan@lse.ac.uk

Adam Osman
Department of Economics
University of Illinois at Urbana-Champaign
109 David Kinley Hall, 1407 W. Gregory
Urbana, IL 61801
aosman@illinois.edu

Dean Karlan
Kellogg Global Hub
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and CEPR
and also NBER
dean.karlan@gmail.com

1 Introduction

When credit markets are characterized by asymmetric information, relational contracts dominate and a specific lender can become the defacto monopoly provider of credit to a set of firms wishing to access larger loans (Ghosh and Ray, 2016). In this circumstance, an individual lender's credit allocation choices can become important determinants of the efficiency of capital allocation, and hence aggregate productivity (Hsieh and Klenow, 2009). Further, if lender credit allocation is inefficient, then the average large loan will generate more default and less firm profit than it could, and equilibrium would be characterized by both misallocation and smaller than optimal loan sizes.

Despite the potential importance, little is known about the quality of lender decisions beyond that which can be inferred from aggregate exercises such as those presented in Hsieh and Klenow (2009) and Hsieh and Klenow (2014).¹ This dearth of knowledge likely stems from three difficulties an analyst faces when attempting to understand the quality of lender credit allocations. First, if poor allocation of credit means large loans have high default rates or are unprofitable for borrowers, equilibrium loans may be too small to reveal heterogeneity. Second, to understand the quality of allocation decisions, one must collect appropriate data to estimate heterogeneous treatment effects and use methods that avoid over-fitting. Finally, there needs to be a credible identification strategy, one that generates both a sufficient quantity of loans to firms that would not normally be allocated credit, as well as determines which firms would have received loans in a business-as-usual loan size expansion.

We collaborated with the Alexandria Business Association (ABA), a large Egyptian lender, to design and implement an experiment that would overcome the above difficulties. ABA's senior management understood that there was demand for larger loans but was concerned about risk, both for themselves and their clients. Our collaboration was designed to understand whether larger loans could be beneficial, and whether credit allocation decisions could help to mitigate risk.

ABA first worked to select a sample of firms that it believed might benefit from a larger loan. They then conducted a randomized experiment within this sample, in which treatment borrowers were offered loans four-times larger than their previous loans, and control borrowers were offered loans that were two-times larger.² The ex-

¹Recent work has focused on a related question, whether loan officers can help choose firms that are likelier to repay and more profitable for the bank, as well as whether loan officer recommendations reveal bias (e.g., Dobbie et al. forthcoming, Fisman et al. 2017 and Cole et al. 2015). Particularly closely related to our motivating questions is Rigol and Roth (2021), which shows that contractual incentives for loan officer can easily be misaligned, even against lender profits. Specifically they show that loan officers incentivized to reduce average default in their client set do not recommend their lowest default clients to receive larger "graduation loans", which are provided by a separate department of the same lender. None of these papers considers the profitability of the borrowing firm, which is our focus and what matters for aggregate productivity.

²Standard operations provided borrowers the ability to borrow at 1.5x their previous loan. The decision

periment meets all the challenges discussed above. Large loans were offered even to firms that would not normally qualify, and the size of loans offered to the large loan group extends well beyond normal lending practice for ABA. To measure heterogeneous treatment effects we collected novel psychometric data at baseline. These data were designed to capture characteristics of the owner relevant to entrepreneurial success. We combine these data with modern machine learning methods from [Chernozhukov et al. \(2020\)](#) to identify group average treatment effects for four quartiles. We call those in the top quartile “top-performers” and those in the bottom quartile “poor-performers”.

Having identified top and poor-performers we use two strategies to understand whether larger loans would be allocated to top-performers in a business as usual expansion. First, we believe a reasonable theory of practice is that lenders target capital expansion toward firms that demonstrate success with small loans, believing these entrepreneurs to be of a type that will be consistently successful. This theory suggests preferential access to large loans would be given to firms that see the largest pre-post change in profits when assigned to the control group. We examine whether top-performers are more or less likely to fall into this category. Second, like many lenders, ABA relies heavily on loan officer opinions to determine credit allocation, and loan officers are strongly incentivized to avoid default. We asked a subset of loan officers at baseline to assess whether individual firms would see an increase in default if given a larger loan, and we look at whether loan officers perceive top-performers as more or less likely to increase their default rate.

Our results reveal the potential importance of credit allocation, even *within* the set of ABA borrowers. Specifically, we find mostly null average impacts of the larger loans, but important heterogeneity in treatment effects, along with strong evidence of misallocation. Entrepreneurs in the top quartile of treatment effects for profits, i.e., the top-performers, increase profits by 55% (se=21%) of the control group mean. Poor-performers, on the other hand, are harmed, experiencing a 52% (se=23%) profit reduction. The same pattern holds for related and downstream outcomes. Top-performers increase their wage bill (122%, se=27%), productivity (6.3%, se=1.1%), and household expenditures (44%, se=10%). Poor-performers on the other hand lose 2.3 (se=0.6) employees and greatly reduce their wage bill. To gauge the importance of these magnitudes we compare an intentional lending strategy that targets likely high performers rather than one that lends to the full set of firms identified as viable by the lender for larger loans. Targeted lending would improve the average treatment effect for pro-

to provide the control group with 2x loans was made by the bank at the suggestion of loan officers because they believed this would make the randomization seem fairer. This choice has some advantages, for instance by keeping even control borrowers engaged with the lender and research team. Borrowers generally had flexibility in determining their loan term, and many in the treatment opted to pay back the larger amount over a longer time-frame.

ductivity from zero to about one-half of a standard deviation, and the treatment effect on monthly profits increases by about 46 percentage points. Thus on 1000 firms, this would increase aggregate monthly profit by about 7 million EGP ($\sim 400\,000$ USD).

However, the top-performers are among the firms least likely to be offered large loans. First, relative to poor-performers, top-performers see a smaller increase (or larger decrease) in profits between baseline and endline when they are placed in control, and this difference is statistically significant. This suggests that loan officers would be unlikely to see top-performers as successful firms, and hence unlikely to recommend them for further loan expansion. Second, directly measured loan officer beliefs strongly militate against lending to top-performers. While loan officers believe that large loans will increase the chance of default by only 22% for poor-performers, this increases to nearly 50% for top-performers. Yet we find no evidence that top-performers are more likely to default.³

Two points are worth noting about the heterogeneity results. First, using only “standard” data of the type that might be routinely collected by a lender (e.g. business performance & demographics), we do not discover any statistically significant heterogeneity.⁴ The heterogeneous treatment effects reported above are identified using a combination of psychometric and cognitive data.⁵ This suggests that (a) a lender seeking to better allocate larger loans will need to collect more data than usual, and (b) the characteristics of the *entrepreneur*, rather than the characteristics of the *firm*, are important for determining the impact of increased access to credit for small firms. While important details need to be worked out – for example, understanding the extent to which psychometric testing is manipulable – our results suggest the *possibility* of improving allocations through the use of better data collection.

Second, while poor-performers perform *relatively* poorly with large loans, they actually have higher *relative* profits in cross-sectional analysis amongst those getting smaller loans. At first blush, this reversal of fortune may seem counter-intuitive; it is contrary to the usual model in which entrepreneurs are credit constrained and high-ability entrepreneurs squeeze relatively more out of scarce resources, whether at low or higher levels of capital. However, we believe that the reversal of fortune is a straightforward implication of two hypotheses that are reasonably uncontroversial. First, a large literature in psychology suggests that there is substantial heterogeneity in individuals’ optimism levels, with a sizable proportion of the population being unrealistically optimistic (Carver et al., 2010, Peterson, 2000, Weinstein and Klein, 1996).

³Full loan default is very rare, since failure to repay loan could result in a prison term. Typically, the lender only pursues legal avenues in cases where they suspect fraud (rather than just low business outcomes). For our study, the lender agreed to not pursue prison terms for anyone in the sample. Borrowers do, however, often deviate from the precise repayment schedule, and it can be costly to ABA to engage in additional monitoring and enforcement activities.

⁴For a more detailed description of both the standard and psychometric data, see section 2.3 below.

⁵Cognitive data includes measures such as Raven’s matrices and also risk aversion measures.

Second, while the finance literature widely accepts that there is a risk reward trade-off it's unlikely that expected gains can be increased indefinitely by taking greater and greater risk, i.e. , the expected income gains that come from taking greater risk ought to be subject to diminishing returns.

We provide a simple model in section 6 that shows how these two assumptions can combine to explain our results. The argument is straightforward. We think of our poor-performers as being over-optimists and our top-performers as being realists. A small loan opens up the opportunity for taking a risky investment and over-optimists take more risk than the realists leading to a larger expected reward. A larger loan, however, opens up even more risk taking opportunities, and induces the over-optimists to increase their risk taking to the point where it actually leads to a reduction in expected reward. No such affect occurs for the less-optimistic, who are better able to appreciate that very high risks are not accompanied by strong returns. These dynamics are in line with previous work that has shown both a sometimes negative relationship between entrepreneur optimism and business outcomes ([Frese and Gielnik, 2014](#), [Hmieleski and Baron, 2009](#)), and a sometimes positive relationship between the two ([Hilary et al., 2016](#)).

We are not able to test the assumption of decreasing returns to risk taking, but we can use our psychometric data to shed light on the hypothesis that the poor-performers are over-optimistic. We do two things. First, we use common machine learning methods to determine which psychometric variables are most predictive of being a poor-performer. This process highlights six psychometric questions that are related to optimism: poor performers respond positively to statements like “in life, failure is not an option,” suggesting optimism about risk taking. Poor-performers are also more likely to agree with statements indicative of a gung-ho attitude, such as “I tend to act first and worry about consequences later” and “when I make decisions I usually go with my first, gut feeling.” Second, we look manually through our set of 50 psychometric questions to isolate those that we think are measuring optimism, and test if they are predictive of being a poor-performer. We find that poor-performers are more likely to agree with statements like “When I make a business decision it is almost always the right decision,” “I have always believed I am going to be successful,” and “I prefer to focus on opportunities rather than risks.” Overall, while not dispositive we believe that these correlations give credence to our optimism based interpretation.

Our paper joins a growing literature that uses different indicators to document heterogeneous firm level returns to capital. [Hussam et al. \(Forthcoming\)](#) finds that individuals are able to identify which peers have the highest returns to capital. [Beaman et al. \(2020\)](#) finds that farmers that borrow have higher returns to capital than those do not. [Banerjee et al. \(2019\)](#) shows how experienced entrepreneurs have much higher returns to credit, while ([Meager, 2020](#)) shows how the returns to credit can be

very positive for some, but also negative for others. [Crépon et al. \(2020\)](#) shows that there is substantial heterogeneity in the returns to both cash drops and loans, and that this may help to explain the difference in impacts between cash drop experiments in the spirit of [De Mel et al. \(2009\)](#) and the more recent microcredit experiments. Also related is [McKenzie and Sansone \(2019\)](#), which finds that neither experts nor machine learning methods predict future entrepreneurial success using “standard” loan application and business data. This question differs from ours in that they are predicting future enterprise success not treatment effects, and they have more limited psychometric data.

We also contribute to the literature on randomized evaluations of credit interventions. The majority of studies in this literature look at the impact of expanding access to micro-size loans, and find modest effects (see [Banerjee et al. 2015](#), [Meager 2019](#)). In contrast to these studies, we study a different margin of credit expansion – access to much larger loans⁶ – and we concentrate on the importance of heterogeneity rather than average effects. Our work is also closely related to papers that look at alternative margins on which loan access can be altered in developing countries. In recent important work [Breza and Kinnan \(2021\)](#) studies the impact of the cancelling of all microcredit access in Andhra Pradesh due to the Krishna loan crisis. They find large impacts on household consumption. The margin on which this experiment alters access is, like our study, very different from the norm: cancelling borrowing for all, rather than marginal borrowers; removing credit rather than adding credit; and general-equilibrium rather than partial. Also related is [Banerjee and Duflo \(2014\)](#) which studies an Indian program that expanded preferential credit access to larger firms (with more than about 150,000 USD in capital stock). This expansion led to substantial investment and increased profits. Finally, [Bari et al. \(2021\)](#) examines providing in-kind loans to existing businesses, which, like our treatment, are four times the size of previous loans. They find statistically significant and positive average effects and do not concentrate on the allocation decision on which we focus here.

2 Study Setting and Sample Characteristics

2.1 Implementing Partner

We partnered with Alexandria Business Association (ABA), a nonprofit organization and the largest microfinance institution in the Middle East by number of clients. At the time of the study ABA had over 400,000 borrowers and fewer than 1% of those

⁶Appendix Table 1 compares loan sizes in our setting to those in the seven studies included in the [Meager \(2019\)](#) meta-study: Using PPP exchange rates, loans in our treatment group range from 2.5 to 15 times larger than the loan sizes in those studies.

borrowers had loans over 1,000USD\13,000EGP.⁷ Senior management expressed interest in expanding the size of loans offered to small and medium enterprises (SMEs). Immediately prior to our project, ABA policy allowed for new loans to increase by up to 50% after successful repayment of a prior loan. Loan size expansion ceases when the firm stops demanding more credit, the loan officer deems the loan to be too risky, or they reach 150,000EGP, which was the the regulatory maximum for MFI's in Egypt at the time of the study.⁸

At the time of our study, ABA provided both individual and group loans and our intervention focused on the individual lending side of the portfolio. Individual loans began at 115USD\1,500EGP and had a ceiling 100,000EGP. Annual interest rates varied slightly depending on market conditions, but ranged from 14% to 17%.⁹ Borrowers could take loans of various maturities, but the majority of loans were provided on 12 month terms. Borrowers could go to a branch office to apply for a loan, but many loans were initiated directly with loan officers who often visit businesses to recruit new borrowers. The application process required individuals to commit to repayment and to include a guarantor.¹⁰ ABA provided both household and enterprise loans, but loans over 20,000EGP were provided only to registered businesses. This requirement was relaxed for our study, allowing many unregistered enterprises to get loans larger than 20,000EGP.

In Egypt, as in many developing countries, there is a “missing middle” of credit availability that may contribute to the lackluster growth of small firms: MFI's typically offer loans up to a maximum of about EGP 5,000 (~USD 280) whereas banks typically start loan sizes at EGP50,000 (~USD 2,800) and have formality requirements that small enterprises often cannot meet. We conjectured that this gap in the lending market could, in part, be attributed to the combination of asymmetric information, which leads to relationship based lending starting from when firms are micro-sized, and poor credit allocation decisions reducing the profitability of larger loans. [Rigol and Roth \(2021\)](#) provides a related explanation.

⁷The average exchange rate during the disbursement of the loans was approximately 13EGP per 1 USD. The PPP conversion factor was about 2.9 during the study period. We use the exchange rate for all conversions unless stated otherwise.

⁸Borrowers could request a larger than 50% increase, but that required approval from the deputy CEO and COO, and only happened in extremely rare cases.

⁹In this context many borrowers may be concerned about the Islamic prohibition on paying interest, but a ruling by Egypt's main religious authority, Al Azhar, states that interest rates charged by non-profit lenders are not prohibited.

¹⁰The guarantor is a person who would be legally liable for the debt in the case of default.

2.2 Experimental Design

Existing clients who were interested and deemed eligible applied for a loan at least four times the size of their previous loan.¹¹ The loans were implemented under a program called *Tamouh*, which translates to “ambition”. To form an eligible cohort of borrowers, the bank first informed loan officers about the lending program and asked them to suggest clients who would be appropriate for a much larger loan. Loan officers then communicated the details of the program to the identified clients, telling them that if their application was approved they would be randomly allocated to either a 4x or 2x larger loan. It is important to note that this design implies the control group was a small deviation from prior policy, in which loan sizes increased by at most 50%, except in rare cases. The decision to provide the control group with 2x loans, rather than the existing policy of 1.5x loans, was made by the bank at the suggestion of loan officers who believed that this would make the randomization seem fairer. Since the selection into the experiment is endogenous, we cannot cleanly interpret a comparison of the 1.5x existing policy to the 2x in the control group.

Part of the capital for the Tamouh loans was provided by Silatech, a regional philanthropic foundation, who agreed to cover some of the risk of default. The partial guarantee was negotiated because of ABA’s concerns that the larger loans would lead to default. This fear of default is common among small and medium enterprise lenders we have interacted with, and part of the purpose of this project was to understand whether this fear is well founded, and if so whether better credit allocation choices could alleviate risk. The existence of the guarantee was known to the Board of Directors of ABA but was intentionally kept from loan officers to avoid any change in enforcement effort. At the time of the experiment, the majority of loan officer compensation was an incentive payment that depended on the number of loans, not the size of loans, and also included a repayment condition. If a loan officer’s portfolio repayment rate was below 97%, they did not receive any incentive payment; for repayment rates above 97% but at or below 98.5% they received 50% of their incentive payment, and for repayment rates above 98.5% they received their full incentive payment. Because repayment rates were based on default relative to overall portfolio size, loan officers expressed concern that the Tamouh loans could harm their overall compensation. To appease this concern senior management added an additional incentive for loan officers, which provided a fixed payment for each client that ended up in the experiment. The fact that ABA felt the need to make this payment reveals the importance of loan officer beliefs in determining credit allocations. Ultimately, all of the loans in the experiment, both treatment and control, were repaid, and thus had no adverse effect on

¹¹ Although the plan had been to allow treatment loans to be up to ten times the size of the previous loan, in practice ABA let treatment loans be typically four times larger but not more

the loan officer's portfolio repayment rates.

To apply, invited borrowers filled out a standard loan application as well as an in-person baseline research survey (conducted by ABA employees from the non-financial services division).¹² ABA surveyors informed borrowers about the goals of the survey and that their data would neither be shared with ABA's financial arm nor affect ABA's lending decision in any way. The baseline survey included detailed questions about the borrower's businesses, demographic questions, and non-standard types of data such as cognitive and psychometric questions. Afterwards, a central credit committee reviewed the borrower's application (without access to the baseline survey answers) and made a lending decision. In all cases the credit committee required that borrowers had successfully repaid at least three prior loans. Approved loans were grouped into batches for randomization. For the randomization, we worked closely with ABA to randomize 31 batches of applicants over 13 months. For each batch, ABA would send the research team a list of between 1 and 133 applicants. For each applicant, we were given the full application material but only used the identity of the loan officer for the purpose of stratification. To stratify, we employed a simple block randomization routine by loan officer. When there was an odd number of applicants with an individual loan officer, we assigned the additional applicant with an independent 50% probability to treatment or control. Our final sample consists of 1,004 borrowers.

Aside from the loan size, the loans adhered to ABA's normal loan terms, monthly repayments, and monitoring and enforcement protocols (more detailed description below, under description of the first stage results).

2.3 Data Collection

We use two data sources – in-person surveys, and administrative data on loan performance. We implemented three rounds of in-person surveys: a baseline implemented before firms were randomized into treatment and control groups; a first follow-up implemented on average 20 months after disbursement of the loans; and, a second follow-up implemented on average 30 months after the disbursement of the loans. The in-person surveys collected two types of data, (1) standard data on borrower characteristics and the business, and (2) cognitive data, risk preference data, and “psychometric data” collected through a set of 50 statements meant to characterize borrower personality traits.

¹²The data collection had to be implemented by employees of the organization due to data collection regulations in Egypt at the time of the study. They were accompanied by trained supervisors from Silatech.

Standard Data on Business and Borrower Characteristics

All three surveys included standard information on business and borrower characteristics. For business questions, we included profits, revenue, expenditures, and the number of employees and their wages. For borrower characteristics, we included age, experience, and education. We also measured household expenditure using a survey question which asked for total expenditure on a set of commonly consumed food items with a one week recall, a second question asking total expenditure on a set of less regular expenditures, such as water, electricity, and transportation, on a one month recall, and finally a question asking total expenditure on irregular purchases like clothes and mobile phones over the past year. We use baseline versions of these measures to check for balance across treatment and control, and as inputs in the machine learning algorithms. We use endline measures of these variables as our primary outcomes.

Psychometric, Cognitive Data, and Risk Preferences

We included 50 “psychometric” questions in our baseline survey. These data take the form of statements that respondents were supposed to assess on a scale from 1 to 5, with a “1” indicating that they disagree strongly with the statement and a “5” indicating that they agree strongly with the statement. The statements are meant measure which psychological traits are associated with the respondent’s personality. These include traits considered to be in the “Big 5” in the psychology literature, such as conscientiousness, extroversion and agreeableness ([Digman, 1990](#)). Examples of the statements include: “I am good at making last minute changes to plans” which is associated with conscientiousness; and “Before I go to sleep I think ‘what could I have done better today?’.” We also included statements that were shown to be associated with people’s proclivity for entrepreneurship. These statements include “In my group of friends I am the most creative person,” and “I have a strong desire to be successful in life” ([Ahmetoglu et al. 2011, 2017](#)). The 50 statements at baseline were a subset of a 140 statement list that was produced for this study. The remaining statements were included in the follow up surveys. The statements were split across the three surveys in an ad hoc fashion given survey time constraints and our lack of a clear theory to guide choices.¹³ We measure cognitive ability during the baseline survey using Raven’s matrices, digit span recall, and a set of questions testing financial literacy knowledge. We also ask questions at baseline to assess risk aversion asking both about overall perception of risk taking, and a question about a hypothetical risky investment.

¹³We included the questions in follow-up surveys with the hope that the psychometric measures would be time-invariant and treatment-invariant and thus enable us to treat them all as “baseline” measures. To test for this, we included some questions at both baseline and follow-up, but we find enough changes over time to eliminate the ability to use the questions asked at follow-up in the heterogeneity analysis.

A full list of the risk, financial literacy, psychometric, and cognitive is included in the Survey Questions section of the Online Appendix.

Administrative Data & Loan Officer Perceptions

Administrative data from the lender provide a detailed account of every payment made by the borrower, including the amount of the payment, the date of the payment, proportion of the payment that went to principal, interest, and any associated late fees. These data cover the five loans prior to the experimental loan as well as any loans up to 30 months after the experimental period.

We also implemented a loan officer survey. The survey collected demographic information about the loan officers as well as information about the loan officers' expectations regarding business and loan performance for the borrowers in the case where the borrowers were allocated to treatment or to control. We asked loan officers to grade borrowers on their "ability to repay the loan" on a 1-10 scale. We asked them the question twice, once about their perception of repayment ability in the case that the borrower received the larger loan, and again in the case that the borrower received the smaller loan. In our analysis, we score these as a binary variables, that combine the two responses, equal to one if the loan officer perceives the larger loan will be more likely to lead to default than the smaller loan. We have similar data for loan officer predictions on revenue increases due to the loans. Due to logistical constraints we were only able to collect these data before randomization for about a third of the sample (293/1004 borrowers), and we do not use the predictions from loan officers whose clients had already received a loan. On average, loan officers expect the likelihood of repayment to decrease with the larger loan relative to the smaller loan for 31% of borrowers and for 32% of borrowers to have a greater ability to increase revenue with the larger loan than with the smaller loan.¹⁴

Sample Characteristics

Table 1 reports sample summary statistics from the baseline survey and balance verification for the randomization. Borrowers are on average 40 years old with 9 years of education; about 20% of the sample are women. For the businesses, about a third are registered, the average tenure is 12 years, and the average number of employees is 0.65 (37% have any employees, and, for those that have any employees, the average number is 1.75). Monthly profits average about 6,800 EGP, equivalent to USD 525. The average size of firm's previous loan from ABA was 7,540 EGP. At the time of the base-

¹⁴If the survey was conducted after randomization and loan provision, we find that treatment assignment affects loan officer perceptions; thus we only use the loan officer surveys that were completed prior to randomization and loan provision.

line survey 6% had borrowings from other sources, with an average outside debt of about 2,000 EGP.

Column 2 reports the baseline difference between the treatment group and control group, after controlling for the stratification variables (loan officer fixed effects). None of the differences is statistically significant, and a joint test of significance, reported at the bottom of the table, yields a p -value of 0.79.

Appendix Table 2 reports summary statistics and balance verification for a subset of the psychometric and cognitive variables. The table reports on six key psychometric variables that correlate strongly with treatment effects (we discuss below how we identify these six questions) as well as the cognitive and risk aversion questions. We find no statistically significant differences between treatment and control groups. In a joint test, we find that the entire set of 50 psychometric variables collected at baseline as well as the cognitive and risk questions do not predict treatment status (p -value=0.53).

We were successful in reaching most of our sample for the 20-month survey (96.3% response rate) and the 30-month survey (95.1% response rate). Column 1 in Appendix Table 3 shows that there is no evidence of differential attrition when we regress a binary for attrition on the treatment indicator directly. Columns 2-5 regress attrition on a large number of baseline covariates (columns 2 & 4) and those covariates interacted with treatment (columns 3 & 5). A joint test of significance on the coefficients of the interacted characteristics produces a p -value of 0.99, providing confidence that differential attrition is not an issue for our analysis.

3 Average Impacts: Intent to Treat Estimates

We use a standard ANCOVA regression framework to estimate average intent-to-treat effects, pooling our two rounds of follow-up data. Specifically, we estimate

$$Y_{it} = \beta_1 T_i + \beta_2 Y_{i0} + \delta_{LO} + \lambda_t + \varepsilon_{it},$$

where we have up to two observations per person (one for each of the two follow-up surveys), Y_{it} is the outcome of interest at time t and β_1 is the coefficient of interest on a binary variable that is equal to 1 for individuals assigned to the treatment group. We include the baseline value of the dependent variable as a control when available as Y_{i0} , loan officer fixed effects as δ_{LO} , and an indicator variable for the second survey round as λ_t . We include loan officer fixed effects since our randomization stratifies at that level, and we cluster standard errors at the individual level.

3.1 Borrowing at ABA (First Stage)

Table 2 reports how borrowing differed by treatment assignment. 76% of the control claimed their loans, while 85% of treatment did so. As intended, we find that the average loan size in the treatment group is more than double the loan size in the control group. Standard loan terms were 12 months, but borrowers were allowed to request longer to repay and requests were considered by the central credit committee. Treatment group borrower's loan terms were 50% longer on average (19.7 months versus 13.2 months on average). Thus while monthly payments for the treatment group were larger than for the control group, the gap is 40% rather than 100%.

Figure 1 shows how outstanding debt from the lender differs across treatment and control groups over time, extending three years after dispersal of the treatment loans. It shows how, in the year preceding the experiment, treatment and control had very similar levels of debt. After the experiment, outstanding debt climbs for the control group but climbs dramatically higher for the treatment group. Given the differences in the length of the loan, the figure shows how the control group's outstanding debt falls around the 12 month mark as their experimental loan is repaid and then increases again, showing another small dip at 24 months. For the treatment group, outstanding debt is at its lowest level at around 22 months after disbursement, around when they pay back their experimental loan, and then shoots back up directly afterwards.

We consider what predicts take-up in Columns 4 & 5 of Table 1. We run a single fully interacted regression where we regress take-up on baseline variables interacted with control and an identical set interacted with treatment. In control, women are more likely to take up the loan, as are those with a registered business, higher monthly profits and lower spending on labor. In treatment, women continue to take up the loan more often, as do people with some employees. Column 6 reports the p-value for equality of the coefficients across treatment and control. Appendix Table 2 repeats this analysis but with the cognitive and psychometric measures and finds limited evidence of any differences across treatment and control groups.

3.2 Average Impacts on Loan Repayment and MFI Profits

Using administrative data from the lender we are able to assess how borrower repayment behavior differed between the treatment and control groups. Table 3 reports that every individual eventually repaid their loan, implying that the concerns over higher default rates were far from realized. Despite full repayment many borrowers struggled to make all of their payments on time, with treatment firms faring worse. 76% of the control group had a perfect repayment record, while only 63% of the treatment group did. Similarly, the average number of days with a payment overdue was 12.5

for control group and 26.7 for treatment group.¹⁵

The increase in late payments naturally leads to an increase in payment penalties, which allows us to assess the impact of these larger loans on the lender's bottom line. Table 3 calculates total penalties paid by borrowers in the first 24 months after disbursement. This timeframe is long enough to account for the fact that some control firms may have taken two loans in the time a treatment firm had only one. Over those two years control firms average 128 EGP in penalties while those in treatment average 323 EGP, an increase of 195 EGP. This increase in penalties paid is much greater than the opportunity cost associated with the money that remains outside the lender during the time the borrowers are late. As indicated in the final row of Table 3, the treatment group on average has 29 EGP more outstanding on any given day. This is money that the lender could not lend out. Multiplying this number by a daily interest rate gives the total foregone revenue per day for ABA from loans outstanding. Even if the daily interest rate was 0.5% (which is far too high) that would mean only 0.15 EGP per day, or a total of 106 EGP over 24 months. In contrast, the total late fees collected by ABA over 24 months was 195 EGP higher in the treatment group, indicating that ABA earned more in late fees than it lost in foregone lending revenue.

To determine whether or not the larger loans were more profitable from the lender's side we also need to consider the potential increased costs of loan officer time dedicated to following up with the borrowers to ensure eventual payment. Loans officers and senior management report that loan officer time did not change significantly due to the offsetting forces of decreased administrative load from fewer larger loans, and additional time spent focusing on following up with the borrowers of the larger loan. Both loan officers and senior management grew to view the program as a success from a business perspective and they expanded it throughout the entire organization after the completion of the experiment. Hence, we conclude that the larger loans were profitable for ABA.

3.3 Average Impacts on Primary Outcomes

We report average impact of treatment on six primary outcomes in Table 4, Column 2, Panels A and B. Profits (Panel A) increase by 1,448 EGP a month, a 9% increase (se=7%) relative to control, thus positive but not statistically significant. Impacts on revenues and expenses are similar, with an increase of 5,715 EGP (12.5% increase, se=4,415 EGP) & 5,251 EGP (18.6% increase, se=3,531) respectively, neither of which is statistically

¹⁵ As discussed above, the treatment group had loans that had longer terms of maturity on average and so we also consider the average number of late days per month of the loan and find that it is 47% higher in treatment relative to control, p-value < 0.01.

significant at conventional levels.¹⁶ We also break out expenditure on employment in particular (“Wage Bill”) and do not find a statistically significant increase (144 EGP, $se=247$, a 6.9% increase relative to control). We also generate a measure of total factor productivity (TFP) utilizing endline revenues, and find an increase of 0.04 standard deviations which is also not statistically significant.¹⁷

Our final primary outcome is household expenditure. There we find an increase of 446 EGP ($se=211$), a 9.3% increase relative to the control group, $p\text{-value} < 5\%$. All together these results provide suggestive evidence that the treatment led to increases in business performance on average, and stronger evidence that the treatment is changing household outcomes.

Table 4 Column 2 Panel C reports average impacts of the treatment on seven additional outcomes. Three are important business outcomes: whether the borrower still has a business, the total number of employees, and the value of assets. We find no statistically significant average impacts across any of these variables. We also consider two downstream measures of health, an index of mental health questions and a self-report on physical health and again find null average impacts. We also include two measures related to loan performance. We reproduce the impacts on total late fees in the 24 months after the experimental loan, which increases on average. We also consider whether or not individuals have an outstanding ABA loan at 30 months after the experimental loan and find a decrease of 8 percentage points in the treatment group.

4 Heterogeneous Treatment Effects

While the average treatment effects suggest a small positive effect of larger loans on business outcomes, it is possible that these average impacts hide important heterogeneity and hence that better allocation of credit could lead to larger average treatment effects. We collected the psychometric data to explore this possibility, given previous work that suggests that these types of data could be predictive of performance (Klinger et al., 2013, Arráiz et al., 2017). To assess heterogeneity we follow Chernozhukov et al. (2020), which provides a strategy for determining whether there is significant treatment effect heterogeneity in randomized experiments where the set of covariates is high dimensional and there is no clear ex-ante hypothesis on which to base a pre-analysis plan.

Because the main aim of the experiment was to increase the profitability of firms,

¹⁶Our profit measure asks business owners to directly report their aggregate profits, instead of asking expenditures and revenue separately and subtracting. This has been shown to be preferable due to the potential for revenues and expenditures to be timed differently in a given month (De Mel et al., 2009).

¹⁷We estimate TFP using a simple specification where we take the residual of a regression of log revenue on log assets and log wagebill and then standardize it. Atkin et al. (2019) shows that revenue based measures of productivity (TFPR) perform better than quantity based measures (TFPQ).

we first look at heterogeneity in the average treatment effect (ATE) for profits; in doing so we follow [Chernozhukov et al. \(2020\)](#), closely and we give a brief description of the method for completeness. We follow a slightly different approach for other outcomes, which we discuss below.

To estimate heterogeneity in the treatment effect for profits, we first split the sample into a “training set” and a “testing set”. Throughout we use 50/50 splits. Using the training set only, we train a machine learning (ML) method to generate a “control” effect $B(Z_i)$ (i.e. the expected outcome for firms with covariates Z if they were assigned to control) and predicted treatment effect $S(Z_i)$, where Z_i denotes the full set of covariates used to predict heterogeneity for subject i . Below we show results where we alter the set of covariates used. Any ML method could be used, but we use four options (elastic net, neural net, random forest, and gradient boosting) and then take the one with the highest prediction score.¹⁸ Note that because we utilize all four ML methods and choose the one with the highest prediction score we utilize a conservative Bonferroni correction in our estimates and multiply all of the p-values by 4, in line with [Chernozhukov et al. \(2020\)](#). In all cases we use the implementation of these methods from the R package caret.

With the estimates $B(Z_i)$ and $S(Z_i)$ in hand we then undertake two analyses using only data from the testing set. First, we estimate the regression

$$Y_i = \alpha * X_i + \beta_1 * T_i + \beta_2 * T_i * S(Z_i) + \epsilon_i \quad (1)$$

where X_i is a set of covariates that includes $B(Z_i)$ and T_i is an indicator for treatment group.¹⁹ Our primary use for this specification is to test the null hypothesis of no heterogeneity $\beta_2 = 0$.²⁰ Second, we split the testing sample into quartiles of predicted treatment effect using $S(Z_i)$ and estimate the regression

$$Y_i = \alpha * X_i + \sum_{j=1}^4 \gamma_j * T_i * 1(S_i \in I_j) + \eta_i \quad (2)$$

where I_j is the set of firms in the j th quartile.²¹ γ_j measures the “sorted group average treatment effect” (GATES) for each quartile, and is the key measure that we use to understand how treatment effects differ across well defined groups.

¹⁸This is defined as $|\hat{\beta}_2|^2 \widehat{Var}(S(Z))$ where β_2 is defined in equation (1) below.

¹⁹The treatment assignment is included as the treatment binary minus a propensity score associated with treatment assignment. The propensity score is constant due to the randomized treatment assignment. The individual treatment effect $S(Z_i)$ is included as a deviation from its mean.

²⁰ $\beta_2 = 0$ if there is no heterogeneity, or the ML prediction $S(Z_i)$ does not capture that heterogeneity. Hence, this test is of a joint hypothesis, that there is heterogeneity and that the ML methods can detect it using the covariates that we have.

²¹Again, the treatment assignment is included as the treatment binary minus a propensity score associated with treatment assignment.

The key contribution of [Chernozhukov et al. \(2020\)](#) is to show how to get theoretically correct inference for these analyses, and, again, we follow their approach. We repeat the split into training and testing sets 100 times (each with a different randomly chosen split) and run the analyses in (1) and (2) for each split. This process produces estimates of the key parameters β_2 and γ_j for each of the 100 splits, as well as the associated confidence intervals, standard errors and p -values. For the parameter estimates we report the *median* from the 100 runs. For a $1 - \alpha$ confidence interval we report the median of each boundary of a $1 - \alpha * 2$ confidence interval from each split. For hypothesis tests in equation (1), we state that a hypothesis is significant at the α level if the median p -value is less than $\alpha/2$. The use of $\alpha/2$ in the hypothesis tests and confidence intervals corrects for sample splitting. As mentioned above, due to the initial test of 4 machine learning prediction methods we implement a Bonferroni correction by multiplying p -values by four.²²

Results are presented in Figure 2 and Table 4, Panel A, Columns (3) - (6). Figure 2 shows GATES for quartiles using three different definitions of Z_i : just the psychometric/cognitive data; just the standard data; and all data (both psychometric/cognitive data and standard data). The first point to note is that there is strong evidence of heterogeneity. Using just the psychometric/cognitive data,²³ we reject the hypothesis of no heterogeneity in treatment effects ($\beta_2 = 0$), and estimate $\beta_2 = 0.85$ with a p -value of 0.002. This coefficient being so close to 1 shows that the model does a very good job at predicting treatment effects accurately.²⁴ Concentrating on the estimates using the psychometric/cognitive data we find in Table 4 Panel A Columns (3) - (6) that those in the most positively impacted group see about a 8,600 EGP increase in monthly profits, and that this impact is significantly different from zero at the 5% level. At the other extreme, those in the least impacted quartile *lose* around 8,200 EGP per month in profits because of the treatment, an effect that is also statistically significant at the 5% level. To put these coefficients in perspective, mean profits in the control group in our sample is about 15,700 EGP per month meaning that the positive gain for the top quartile is about 55% of the control mean, while the loss for the least effected group is about 52% of the control mean.²⁵ As a robustness check we replicate the analysis but replace the profit variable with a random variable that has the same mean and standard deviation as the actual profit variable. In that case we fail to reject the null hypothesis, with an

²²We note that this procedure is quite conservative.

²³When we refer to “psychometric/cognitive data” we include all 50 psychometric measures, as well as our measures of risk aversion and cognitive ability. When we implement the analysis without the risk aversion or cognitive ability measures our estimates of heterogeneity are even stronger.

²⁴We also find statistically significant evidence of heterogeneity using other machine learning algorithms included in the [Chernozhukov et al. \(2020\)](#) method, but we focus on the estimates from the method with the highest prediction performance.

²⁵This increase in profits suggests a monthly rate of return of nearly 11% on the additional capital for top-performers, about twice the average rate of return found in [De Mel et al. \(2009\)](#). Given that these returns are for those who are most capable of utilizing the larger loan this seems reasonable.

estimated β_2 of 0.11 and a p -value of 0.80, and no statistically significant difference between the top and bottom groups.²⁶

It is important to note is that the psychometric/cognitive data does a much better job of capturing heterogeneity in treatment effects than does the standard data, and in fact inclusion of the standard data adds noise. When we utilize all of our data the p -value associated with the heterogeneity is 0.237 and the difference in GATES between the top and bottom group is about 6700 EGP. In contrast when we restrict our data only to the psychometric/cognitive questions we rule out homogeneity with a p -value of 0.002 and a bottom vs. top group difference of over 16,000 EGP. Finally, when we use only the standard data the estimated p -value increases to 0.483 (so that we cannot reject the joint hypothesis of zero heterogeneity or that the standard data cannot detect the heterogeneity). This implies that primarily the psychometric data are allowing us to predict the heterogeneous treatment effects. This result also implies that these machine learning techniques cannot fully differentiate signal from noise, as evidenced by the decrease in performance when using all of the data relative to only the psychometric/cognitive data. While the algorithms do help deal with the high dimensionality of the data, there is still an important role for researchers to identify which variables to include, and in what form ([Mullainathan and Spiess, 2017](#)).

These estimates show that there is a subset of the sample who are using the loan to greatly expand their business, increasing revenues and expenditure and spending on labor. These individuals are also using the additional business profits to increase their expenditure at home. On the other side of the spectrum are individuals who were unable to turn the larger loan into business profits, and instead are dealing with a large decrease in the size of their business across revenue, expenditure, and labor outlays, with suggestive evidence that the negative business impacts also extended to the home.

Critically, these groups are not generated in a way that overfits the data to maximize heterogeneity, rather they are produced through a method that utilizes baseline data and split-sample validation. This means that the Individual Treatment Effects (ITEs) estimated for each person do not use any of the data on that individual's performance. Hence, it is not equivalent to grouping individuals who may have realized a negative/positive shock in their business and splitting up the sample based on realized outcomes. Instead, this categorizes people based on their baseline characteristics and the algorithms' predictions about how they would perform with the larger loan based on the performance of other people who are observationally similar at baseline.

The results indicate that there is a "type" of borrower who will be able to grow

²⁶We also replicate the analysis while dropping individuals who did not take out a loan in either treatment or control and find very similar results, suggesting that the observed heterogeneity is not driven by differential loan take up rates.

their business with larger loans and a “type” of borrower who will actually shrink their business when provided a larger loan relative to a smaller loan. This is particularly striking given the several layers of review that were already built into the selection of the sample. In addition to the borrowers requesting the loan on their own behalf, the loan officers needed to approve of the individual getting a larger loan, the borrower had to have successfully completed three previous loans, and a central credit committee needed to review their case before they were randomized. Hence all of the links across that chain were intended to weed out individuals who would not benefit from the loan, but even in this highly selected sample there is significant heterogeneity and a group of borrowers who are negatively impacted. We explore this further in the following section.

These results are in line with earlier work on microcredit that show evidence of important heterogeneity in impacts. For example, [Banerjee et al. \(2019\)](#) finds that microcredit leads “gung ho entrepreneurs” to more than *double* profits relative to those in control. [Crépon et al. \(2020\)](#) also provides evidence that credit can double profits, and is particularly heterogeneous as well. Related, [Meager \(2020\)](#), in a seven-site meta-study employing Bayesian hierarchical modeling, finds large positive impacts for a subset of borrowers *and* evidence of negative returns to credit for some borrowers in two of the seven sites (Mongolia & Ethiopia).

4.1 Conditional Group Average Treatment Effects

Having found heterogeneity with respect to profits, in this section we explore whether firms that saw an increase in profits also saw changes in other related outcomes. Our aim is to shed some light on mechanisms and to validate the plausibility of our results. For example, understanding whether the top performers, those with the highest ITEs on *profit*, also increased revenue, and wages, etc., can build insight regarding the likely path towards increased profits. For example, do profits increase via an expansion of the business or via a shrinkage of unprofitable inputs? The former implies a likely increase in wage bill; the latter implies a reduction. While the method from [Chernozhukov et al. \(2020\)](#) allows for assessing heterogeneity on any variable, it does not directly allow for determining whether heterogeneity in profit treatment effects translate into heterogeneity in other variables. Hence, we extend the GATES procedure from [Chernozhukov et al. \(2020\)](#) to estimate Conditional Sorted Group Average Treatment Effects, or “CGATES”.²⁷

The CGATES procedure follows the normal GATES procedure closely. For each of the sample splits (into training and testing) we train our ML method on the training data and calculate $S_i = S(Z_i)$, the predicted individual treatment effect for i , for each

²⁷This analysis is similar what is implemented [Bertrand et al. \(2021\)](#).

individual in the *testing* group. At the end of this procedure, we take the median of these individual treatment effects for each i and rank individuals to create quartile rankings according to the individual treatment effect *for profits*. We then run the same specification as in equation 2 above, but change the Y variable to be the other outcomes variables of interest. In particulate we estimate

$$Y_i = \alpha * X_i + \sum_{j=1}^4 \gamma_j * T_i * 1(S_i \in I_j) + \eta_i \quad (3)$$

where I_j is an indicator for individual i being in the N th quartile ITE group for *profits*, which we refer to as the N th CGATES group.²⁸

We are interested in γ_j , which is the average treatment effect for individuals who were allocated to the four different CGATES groups. The main difference between this method and the [Chernozhukov et al. \(2020\)](#) method is that the group assignment is based on profits and not on the other outcomes of interest that we analyze. This allows us to assess how impacts differ for individuals who increased their profits in response to the larger loan and those who decreased their profits. We also take the median value of S_i from the 100 simulations and then allocate individuals to groups, instead of taking the median coefficient from regressions across the 100 splits of the data. We then apply a Bonferroni correction by multiplying p-values for each coefficient variable by four as above.

Columns (3) - (6) of Panels B and C in Table 4, report the results of these regressions. Panel B shows impacts on our primary outcomes other than profits. Column 3 reports the treatment effect for individuals who were in the first CGATES group (those in the lowest quartile of individual treatment effects for business profits). Columns 4-6 follow the same format, reporting the treatment effect estimates for individuals who were in the second, third, and fourth CGATES groups. The individuals in each group are held constant across all of the regressions in Panels B and C of Table 4 since we are conditioning on the ITE group for profits, and not for the ITE groups optimized for the outcomes presented. Many of these outcomes are highly correlated, and so we expect results to be similar to the impact on profits, but this exercise allows us to better understand the mechanisms through which profits are increasing.

The first row in Panel B shows that those in the highest ITE quartile for profits also have a large increase in revenues, with an increase of 47,556 EGP (se=9481), while those in the lowest CGATES group see a revenue *decrease* of 51,463 EGP (se=10406). A similar pattern holds for the other business outcomes in the subsequent three rows, expenditures, spending on wages, and estimate TFP: there are large statistically sig-

²⁸Again, the treatment assignment is included as the treatment binary minus a propensity score associated with treatment assignment, and the X vector of controls includes the predicted values of $B(\cdot)$ & $S(\cdot)$ for profits.

nificant gains for those in the top CGATES group and losses for those in the bottom CGATES group. For example, those in the top CGATES group increase their wage bill by nearly 2,400 EGP per month ($se=532$), while those in the bottom CGATES group decrease their wage bill by about 2,000 EGP ($se=628$).²⁹ We also find large differences in estimated TFP in the top and bottom quartiles. The final row in Panel B shows that the heterogeneous effects in business outcomes are also reflected in reported household expenditure. The top group experiences a statistically significant 2,077 EGP increase in expenditure, while the bottom group sees a decrease of 291 EGP, although this decrease is not statistically significant.

Panel C continues our analysis of heterogeneous treatment effects showing estimate treatment effects for each CGATES group on additional outcomes. The table shows that for those in the bottom CGATES group the larger loan decreased the likelihood that they were able to keep their business open by 8 percentage points. The loan also led to an increase in the number of employees by 0.84 employees for the top group (although this increase is not statistically significant) and a decrease of 2.27 employees in the bottom group. The point estimates for changes in the value of assets are large but not precisely estimated. Estimates of the heterogeneous impacts on mental and physical health suggest a similar pattern, with increases for those in the top profits group and decreases for those in the bottom groups. It's worth noting that there is a decrease in reported mental health of 0.26 standard deviations for those in the bottom quartile. This decrease is both large and statistically significant and shows the mental toll that comes from a failing business.

We also explore heterogeneity in borrowing behavior. We find some evidence that those in the bottom group pay more in late fees, but this result is not statistically significant. We also find that those in the lowest group are less likely to be borrowing after 30 months, as are those in the third highest quartile, implying some non-linearities in impacts.

5 The Quality of Allocations

The heterogeneity of impacts discussed above suggest substantial scope for the allocation of large loans to affect the impact of credit availability and the aggregate profitability and productivity of firms borrowing from ABA. We now discuss whether, in the absence of psychometric data and machine learning techniques, ABA would be in

²⁹The decrease in spending on labor is larger than the average value in the control group and midline and endline. This is likely due to non-linearities in the size of the control group wage bill in the follow up survey by treatment quartile, those in the highest and lowest quartile have the largest wage bill when in control, and it is greater than the 2,000 EGP reduction for the lowest quartile. It's worth noting that there are no differences in *baseline* values of wage bill spending between any of the quartiles. This is also true when we consider the number of employees.

a position to make efficient allocation decisions. As noted above, to study the allocation problem requires providing large loans to firms that would not normally receive loans in a “business as usual” credit expansion, and as a consequence we do not have a direct measure of which firms would have received loans in the absence of our intervention. We hence use two different proxies that we believe provide a good sense of which firms would likely have been allocated credit.

5.1 Firms that Perform Well with Small Loans

A natural theory of loan allocation is that the lender will perceive firms that perform well with smaller loans as high quality entrepreneurs, and hence will be more likely to extend larger loans to them. This would also be an optimal loan allocation strategy in models such as [Banerjee and Moll \(2010\)](#), which posits a constant firm specific productivity term that multiplies a decreasing returns to scale production function. According to this theory, the allocation of large loans is likely to be relatively efficient if top-performers are among those firms which see the greatest firm growth when given a small loan, and inefficient if top-performers perform relatively badly with small loans. Evaluating the quality of allocation in this way has the advantage that it moves beyond the specific allocation decisions of ABA to evaluate the effectiveness of this simple allocation rule on the firms within our sample, which we believe to be relatively representative of SME’s in Egypt.

Table 5 presents evidence consistent with a propensity toward misallocation of larger loans, if our basic theory is correct. Each column in the table is a regression in which we regress a measure of firm performance on a standardized measure of our predicted individual treatment effect. The top-performers have high values for this index. Panel A has baseline to endline change in profits as the outcome measure of firm performance, while panel B has endline TFP.³⁰ Columns 1 and 2 show that, when they are placed in the control group, our top-performers do badly: they are the firms that see the smallest baseline to endline increase in profits, and also the firms that have the smallest endline TFP. In contrast, columns 3 and 4 show when the top-performers are placed in the treatment group, they see the largest increases in profits and have the highest endline TFP. The availability of credit leads to a strong reversal of fortune. Overall, these results suggest the entrepreneurs who benefit most from the loan gain, because, in the absence of the loan, they would have done poorly. Those who lose from the loan do so because they would have done well without it. It also implies that misallocation will ensue if a lender pursues a strategy of targeting large loans at those firms who perform well with small loans.

³⁰We do not have the data necessary to compute baseline TFP.

5.2 Loan Officer Perceptions

Our second strategy for understanding likely misallocation is more strongly tied to ABA's institutional structure. As noted above, loan officer incentives strongly favor minimizing default, and loan officers have a great deal of control over which firms get loans. This implies that if loan officers believe top-performers will see relatively low defaults then large loans will be well allocated. Simultaneously, it also implies that, if loan officers perceive top-performers to be relatively risky, they will misallocate large loans. In favour of the reasonableness of this assumption, we also asked loan officers to list their most important considerations when deciding whether or not to provide a loan to a client. 70% listed repayment ability as the top reason, with another 20% listing it as the second most important reason.

Panel A in Table 6 presents results on loan officer perceptions using data collected before the randomization.³¹ We asked loan officers to grade borrowers on their "ability to repay the loan" on a 1-10 scale. We asked them the question twice, once about their perception of repayment ability in the case that the borrower received the larger loan, and again in the case that the borrower received the smaller loan. We score this as a binary variable that combines the two responses, equal to one if the loan officer perceives the larger loan will be more likely to lead to default than the smaller loan.³² We compare their responses to these questions for top-performers relative to poor-performers.

We find that loan officers believe that those in the top quartile of impacts are *more* likely to struggle to repay if provided the larger loan relative to those in the bottom quartile. Loan officers predict that a large loan will increase default for 22% of poor-performers and this increases by a statistically significant 24 percentage points for top-performers. This fact strongly suggests that, in the absence of some alternative method to identify top-performers, large loans would be misallocated.³³ Panel C in Table 4 also suggests a disconnect in observed repayment behavior, with those in the bottom quartile paying more in late fees than those in the top quartile, although these estimates are not statistically different.

While ABA makes its lending choices based on perceived default, an alternative way to allocate credit would be for loan officers to encourage loans to those firms that they believe will benefit most from them. To test whether this strategy would improve

³¹ As noted above, due to logistical constraints we were only able to start the loan officer survey towards the end of implementation, and only have loan officer predictions for the final 29% of the sample.

³² When we implement this analysis but using only the loan officer's assessment of performance on the larger loan we find similar results.

³³ Panel A of Appendix Table 4 includes a related analysis. We split the sample by loan officer prediction of the impact of the loan on repayment ability and estimate the impact on profits for each group. We find that loan officer perceptions about repayment ability are negatively correlated with impacts on profits.

allocation we asked loan officers to predict which firms would see a revenue increase. Again Panel A in Table 6 presents results on how these perceptions correlated with treatment effects. We find no statistically significant difference between the top and bottom-performers, although the sign is in favor of efficient allocation: the loan officers believe that 24% of poor-performers and 37% of top-performers will experience more revenue growth with the larger than with the smaller loan. While weak, this evidence suggests that allocation based on perceived revenue gains would be more efficient than current practice, but not as effective as allocation based on psychometric data.³⁴

6 An Explanation From Optimism and Risk Taking

Our main findings, particularly the reversal of fortune documented in Table 6, may seem surprising. The result is inconsistent with many models used to study capital misallocation, which tend to assume that entrepreneurs draw a productivity type z_i which multiplies a DRS production function to give output $z_i f(k_i)$ (e.g., [Banerjee and Moll 2010](#)). In this section we argue that our results are in fact easily generated by combining two reasonable assumptions: (1) decreasing marginal returns to risk taking; and, (2) heterogeneous levels of optimism, with some individuals being over-optimistic. Because we think these assumptions are likely to hold broadly, we believe that our findings are likely to be relevant in a range of settings.

We first motivate the theoretical framework with a stylized example. Consider two entrepreneurs, who each own a restaurant. Omar is an optimist, and Rana is a realist. Both wish to expand their business, and apply for a loan to do so. With a small loan each finds a way to increase profits. Rana uses the loan to replace some kitchen equipment, a low risk cost control option. Omar, on the other hand, buoyed by his optimistic belief that things will work out in his favour, takes a slightly greater risk and expands the size of his existing restaurant. Both increase profits, but Omar increase profits by more.

Next consider a larger loan. The larger loan opens up new possibilities, and both entrepreneurs now have two options: continue expanding their existing restaurant, or open a second location on the other side of town. Rana, being the realist, is concerned about over-expansion and decides to continue growing her original restaurant. With the larger loan she is able to take more actions and sees a larger increase in her profits than with the small loan. Omar, on the other hand, believes that he can handle expansion to a second-location. In reality, the second-location expansion is too risky, a fact that Omar's optimism has caused him to ignore; the second location performs poorly, and the large loan causes a *decrease* in Omar's profits.

³⁴Panel B of Appendix Table 4 again provides related analysis. It splits the sample based on loan officer perceptions and shows that loan officer perceptions are not very predictive of loan impacts on profits.

Omar and Rana's stories are consistent with our empirical findings: first, there is heterogeneity in impacts of the large loan and the large loan causes Omar's profits to decrease; he is a poor performer. Second, we see a reversal of fortune– the optimist Omar does better with the small loan, and the realist Rana does better with the large loan.

Slightly more formally, assume that each entrepreneur in our experiment has a range of projects to choose from. Each project $p \in P$ yields income y_H^p in a "high" state of the world and y_L^p in a "low" state of the world. We assume that the true probability of state H is π_H . The set of projects in P leads to a production possibility frontier (PPF). Line ab in Figure 3 shows what we consider to be a reasonable PPF; it has two key characteristics. First, there is a range of projects over which an increase in risk is rewarded with an increase in expected return.³⁵ These are the modest projects that Rana chose above. Second, and this is our first key assumption, the expected return to risk taking is decreasing in the amount of risk. It seems unlikely that the usual risk return relationship should hold indefinitely: continuing to increase y_H , even at the cost of lowering y_L , is likely to be a difficult task. This argument suggests that at some point y_H can only be increased further at the expense of reducing expected return. As in the story above, a large scale expansion is very risky and may be accompanied by a reduction in expected returns. The concavity of line ab in Figure 3 captures these ideas.

Next, it seems reasonable to assume that the set of available projects depends on loan size. We make two assumptions. First, any project that is feasible with a small loan is feasible with a large loan. This should be uncontroversial and implies that the PPF with a large loan lies everywhere above the PPF with a small loan. Second, the PPF with large loans has a greater (absolute) slope for every y_L . This amounts to assuming that larger loans allow for a greater return to risk taking. In terms of the example above, Omar would surely have expanded to a second location even with a small loan, but that risk would almost surely come with a low return. Over-expanded and strapped for cash, there is likely a very low reward for this extreme level of risk taking. With a larger loan, the situation is probably somewhat better, with each unit of expansion more likely to be rewarded with a positive return. Figure 3 again shows the implications. Curve ab represents a possible PPF for a small loan, while cd is a possible PPF for a large loan.

Finally, we add our second key assumption, that optimism is heterogeneous with some individuals being over-optimistic. We capture this by assuming that there are two types of entrepreneurs: O for optimists and R for realists. We denote the beliefs of optimists π_H^O and the beliefs of realists π_H^R and we assume $\pi_H^O > \pi_H^R$. Second,

³⁵In the figure we assume that $\pi_H = \frac{1}{2}$ so that the line EF , which is perpendicular to the 45 degree line marks an iso-expected return line.

we assume that optimists are “over-optimistic” in the sense that $\pi_H^O > \pi_H$. Both of these assumptions has strong support in the psychology literature (Carver et al., 2010, Peterson, 2000, Weinstein and Klein, 1996, Frese and Gielnik, 2014, Hmielecki and Baron, 2009, Hilary et al., 2016).³⁶

We now wish to demonstrate two potential outcomes in our setting. First, our model can accommodate negative returns to a large loan for those who are optimists. Returning to Figure 3, the lines AB and CD are indifference curves for an optimist, with CD associated with higher utility. These indifference curves are less steep than the iso-expected return line EF , reflecting the optimists belief that the high state is more likely than it actually is. Given these indifference curves, point x is preferred to the point y . But point x has a lower expected return, which can be seen because points x and y lie on opposite sides of the line EF . Just like Omar, the optimist here believes that a good outcome is more likely than it really is and pushes forward with a risky project that a more sober realist would avoid.

Second, our model can explain the reversal of fortune that we see in the experiment. Figure 3 again demonstrates. As noted the curves AB and CD are indifference curves for the optimist. In contrast, the realist’s indifference curves are the dashes lines parallel to the line EF on the 45 degree line. With the small loan PPF, the realist chooses point m , while the optimist chooses point y . The optimist does better with the small loan because his optimism causes him to take greater risk. Over some range greater risk, he is rewarded with greater expected return. With the larger loan the optimist now chooses point x , which as we noted above reduces his expected earnings. In contrast, the realist chooses n increasing her expected return, and also earnings more than the optimist with the large loan.

This simple model is consistent with our experimental findings. While it is hard to test the first key assumption, decreasing returns to risk taking, in the next section we provide some evidence on the optimism or overconfidence assumption.

7 Who Benefits Most from Big Loans?

The previous section argues that our main results, which suggest large heterogeneity in returns to big loans accompanied by a reversal of fortune, can be explained by heterogeneity in levels of optimism. To understand whether this explanation is likely, Table 6 compares the baseline characteristics of top and poor performers. While the machine learning algorithms are non-parametric in nature, comparing average baseline values across groups can help us better understand what differentiates the two groups.

³⁶We could also model optimism as heterogeneity in the belief about the extent of decreasing returns to risk taking. This leads to precisely the same results as presented in the text.

In Panel B we consider standard data collected at baseline, split into demographic characteristics and business characteristics. Those in the top group are less likely to be female and are more educated on average by one year. We find no statistically significant differences in baseline business characteristics including whether the business is registered, baseline profits, revenues, expenditures, or the number of employees. There is also no difference in the size of their previous ABA loan, but those in the top group had about half the amount of loans outstanding with organizations other than ABA, perhaps indicating that those in the top treatment group may have been more credit constrained prior to the experiment.³⁷

In Panel C we consider our baseline psychometric measures, split into cognitive & risk aversion and then the more explicit psychometric measures. We find that those in the top group perform better on the digit span recall task, and on a battery of financial literacy questions. There are no statistically significant differences on their Raven's Matrices scores, and those in the top group exhibit higher levels of risk aversion as evidenced by a lower proportion of investment in a hypothetical risky asset relative to the control group.³⁸

It is less obvious how to compare the top and bottom groups in their responses to our psychometric questions, since we asked 50 questions at baseline and the way these measures are mapped onto specific psychological constructs often relies on researcher discretion. We deal with this challenge in two ways. First, we assess which of the 50 psychometric questions are more predictive of the individual treatment effects estimated in utilizing the [Chernozhukov et al. \(2020\)](#) method. Second, in line with our model about over-optimism, we manually identify additional questions from the 50 that we think speaks to optimism.

We use two methods to determine which of the 50 measures are most predictive of impacts. First, we run a lasso regression with all of the psychometric, risk, and cognitive variables on the right-hand side, predicting the estimated individual treatment effect and chose the top five most predictive questions.³⁹ Second, we build a random forest model of the predicted individual treatment effects using the same baseline variables and recover measures of "feature importance" ([Athey and Imbens, 2019](#)). In essence this measure tells us how often each variable is used when creating the set of

³⁷In Appendix Table 6 we regress the estimated individual treatment effect on baseline covariates and find that previous loan size is positively correlated with it. We include this variable in our "standard data" and it is not sufficient to pick up heterogeneity there.

³⁸We ask individuals to imagine that they had 10,000 EGP that they could invest in an opportunity that would either double their money or halve it, each with a 50% probability. They then report the amount of those funds they would invest in the risky asset.

³⁹Selection of variables in lasso regressions can be sensitive to the inclusion of highly correlated potential explanatory variables. To overcome this limitation we utilize the "Puffer Transformation" as developed in [Jia and Rohe \(2015\)](#) and utilized in [Banerjee et al. \(2021\)](#) to generate stable lasso selections. We utilize cross validation to estimate the appropriate lambda value in the lasso regression, which determines how many variables should be included in the parsimonious model.

decision trees in the model. Each method yields five measures for inclusion, four of which overlap. Thus in Panel C, we report the full set of six that are by-products of these two methods.

Both methods choose the following four statements as important predictors of the individual treatment effect: (i) “I tend to act first and worry about consequences later,” (ii) “I can think of several solutions to any problem,” (iii) “I prefer to have a flexible schedule - I don’t like being tied down,” (iv) “I feel anxious outside my comfort zone.” The Random Forest algorithm additionally chooses the statement (v) “When I make decisions I usually go with my first, gut feeling,” and the Lasso chooses (vi) “In life, failure is not an option.” Across all of these questions those who have high treatment effects are more likely to *disagree* with these statements. We consider these responses to be consistent with our optimism based interpretation of the results. Those who benefit most from the large loans are those who are careful and shrewd, while the poor performers are overly optimistic about themselves and act impetuously.⁴⁰

Next, we manually consider the remaining psychometric statements that we collected at baseline and identified three statements that we feel closely align with optimism. In particular, we look at responses to (i) “When I make a business decision it is almost always the right decision,” (ii) “I have always believed I am going to be successful” & (iii) “I prefer to focus on opportunities rather than risks.” In all three cases poor-performers are more likely to agree with these statements, and hence are more optimistic, in line with the results from our model.

8 Policy Implications & Conclusion

Taken as a whole our results suggest both that the within lender allocation of credit matters for firm profitability and productivity, and that standard approaches to determining how to allocate credit are likely inefficient. The results are striking, and, as with many such results, replication is warranted, ideally across markets and contexts, in order to understand whether the psychometric results are consistently predictive of heterogeneity. Clearly, the underlying characteristics those data are picking up may interact with some contextual factors or loan term feature, thus replication and extensions could help understand these markets and entrepreneurs better. Our work also suggests the potential efficacy of psychometric data in improving allocation. Our slate of psychometric questions are predictive of treatment effects, but are also potentially

⁴⁰Those in the lowest treatment group, the over-optimists, indicate stronger agreement with the several statements, our interpretations are: “I tend to act first ...” & “When I make decisions...” & “I prefer to have a flexible schedule...” indicates impetuosity; “In life, failure is not an option” & “I can think of several solutions ...” hard work, but also a lack of forethought and realism; and “I feel anxious outside my comfort zone” this one seems at odds with our basic theory although, based on the answers to the other questions, over-optimists may have larger comfort zones.

subject to gaming. Individuals were informed that their loan decision would not be based on their answers to these questions, and so our results do not reveal what would happen if lending decisions depended on the responses. Finding strategy-proof ways to detect which individuals would benefit most from large loans is a fruitful avenue for future research.

We also find that many borrowers choose to take out a loan that is larger than what they are able to handle. We interpret this as evidence that some entrepreneurs are overly optimistic and hence more risk-loving, (as evidenced by their with reported ability to “think of several solutions to any problem,” and their greater reported risk taking preferences). This highlights *belief elicitation* as an important gap in the literature on firm development. While some recent and important work has started tackling this (e.g., [Hussam et al. \(Forthcoming\)](#), [McKenzie \(2018\)](#)), much more needs to be learned both regarding how to elicit beliefs (both for researchers, i.e., from survey methods, as well as for practitioners, i.e., for targeting of services) as well as how to incorporate beliefs into our understanding of heterogeneity to returns to policies.

Our results show the important role credit expansion can have on firm growth, but with a strong caution that tremendous heterogeneity can lead to both winners and losers. Many policymakers consider supporting firm growth to be a primary objective, but previous studies have shown that this is often ineffective on average for small loans. We show that a relatively large infusion of credit could help firms grow. Yet, because of the heterogeneous treatment effects, this is not an economy-wide prescription. While some firms benefited from a large expansion in borrowing, others were harmed, even leading to firm exit. Making use of these findings therefore requires either a better understanding of how to improve the allocation of credit, or a belief that expanding larger loans to all firms will let bad firms exit, opening up additional room for the most productive firms to flourish. While plausible, this second line of thinking requires careful consideration of the human costs associated with failed businesses.

References

- AHMETOGLU, G., T. CHAMORRO-PREMUZIC, B. KLINGER, AND T. KARCISKY (2017): *The Wiley Handbook of Entrepreneurship*, Wiley Online Library.
- AHMETOGLU, G., F. LEUTNER, AND T. CHAMORRO-PREMUZIC (2011): "EQ-nomics: Understanding the relationship between individual differences in trait emotional intelligence and entrepreneurship," *Personality and Individual Differences*, 51, 1028–1033.
- ARRÁIZ, I., M. BRUHN, AND R. STUCCHI (2017): "Psychometrics as a tool to improve credit information," *The World Bank Economic Review*, 30, S67–S76.
- ASIEDU, E., D. KARLAN, M. LAMBON-QUAYEFIO, AND C. UDRY (2021): "A call for structured ethics appendices in social science papers," *Proceedings of the National Academy of Sciences*, 118.
- ATHEY, S. AND G. W. IMBENS (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.
- ATKIN, D., A. K. KHANDELWAL, AND A. OSMAN (2019): "Measuring productivity: Lessons from tailored surveys and productivity benchmarking," in *AEA Papers and Proceedings*, vol. 109, 444–49.
- BANERJEE, A., E. BREZA, E. DUFLO, AND C. KINNAN (2019): "Can microfinance unlock a poverty trap for some entrepreneurs?" .
- BANERJEE, A., A. G. CHANDRASEKHAR, S. DALPATH, E. DUFLO, J. FLORETTA, M. O. JACKSON, H. KANNAN, F. N. LOZA, A. SANKAR, A. SCHRIMPF, AND M. SHRESTHA (2021): "Selecting the most effective nudge: Evidence from a large-scale experiment on immunization," Working Paper 28726, National Bureau of Economic Research.
- BANERJEE, A., D. KARLAN, AND J. ZINMAN (2015): "Six randomized evaluations of microcredit: Introduction and further steps," *American Economic Journal: Applied Economics*, 7, 1–21.
- BANERJEE, A. V. AND E. DUFLO (2014): "Do firms want to borrow more? Testing credit constraints using a directed lending program," *Review of Economic Studies*, 81, 572–607.
- BANERJEE, A. V. AND B. MOLL (2010): "Why does misallocation persist?" *American Economic Journal: Macroeconomics*, 2, 189–206.
- BARI, F., K. MALIK, M. MEKI, AND S. QUINN (2021): "Asset-based microfinance for microenterprises: Evidence from Pakistan," *CEPR Discussion Paper No. DP15768*.
- BEAMAN, L., D. KARLAN, B. THUYSBAERT, AND C. UDRY (2020): "Self-selection into credit markets: Evidence from agriculture in Mali," Working Paper 20387, National Bureau of Economic Research.

- BERTRAND, M., B. CRÉPON, A. MARGUERIE, AND P. PREMAND (2021): “Do workfare programs live up to their promises? Experimental evidence from Cote D’Ivoire,” Working Paper 28664, National Bureau of Economic Research.
- BREZA, E. AND C. KINNAN (2021): “Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis,” *The Quarterly Journal of Economics*, 136, 1447–1497.
- CARVER, C. S., M. F. SCHEIER, AND S. C. SEGERSTROM (2010): “Optimism,” *Clinical psychology review*, 30, 879–889.
- CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNÁNDEZ-VAL (2020): “Generic machine learning inference on heterogeneous treatment effects in randomized experiments,” ArXiv: 1712.04802.
- COLE, S., M. KANZ, AND L. KLAPPER (2015): “Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers,” *The Journal of Finance*, 70, 537–575.
- CRÉPON, B., M. EL KOMI, AND A. OSMAN (2020): “Is it who you are or what you get? Comparing the impacts of loans and grants for microenterprise development,” Working Paper.
- DE MEL, S., D. J. MCKENZIE, AND C. WOODRUFF (2009): “Measuring microenterprise profits: Must we ask how the sausage is made?” *Journal of Development Economics*, 88, 19–31.
- DIGMAN, J. M. (1990): “Personality structure: Emergence of the five-factor model,” *Annual review of psychology*, 41, 417–440.
- DOBBIE, W., A. LIBERMAN, D. PARAVISINI, AND V. S. PATHANIA (forthcoming): “Measuring bias in consumer lending,” *Review of Economic Studies*.
- FISMAN, R., D. PARAVISINI, AND V. VIG (2017): “Cultural proximity and loan outcomes,” *American Economic Review*, 107, 457–92.
- FRESE, M. AND M. M. GIELNIK (2014): “The psychology of entrepreneurship,” *Annu. Rev. Organ. Psychol. Organ. Behav.*, 1, 413–438.
- GHOSH, P. AND D. RAY (2016): “Information and enforcement in informal credit markets,” *Economica*, 83, 59–90.
- HILARY, G., C. HSU, B. SEGAL, AND R. WANG (2016): “The bright side of managerial over-optimism,” *Journal of Accounting and Economics*, 62, 46–64.
- HMIELESKI, K. M. AND R. A. BARON (2009): “Entrepreneurs’ optimism and new venture performance: A social cognitive perspective,” *Academy of management Journal*, 52, 473–488.
- HSIEH, C.-T. AND P. J. KLENOW (2009): “Misallocation and manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124, 1403–1448.

- (2014): “The life cycle of plants in India and Mexico,” *The Quarterly Journal of Economics*, 129, 1035–1084.
- HUSSAM, R., N. RIGOL, AND B. ROTH (Forthcoming): “Targeting high ability entrepreneurs using community information: Mechanism design in the field,” *American Economic Review*.
- JAYACHANDRAN, S., M. BIRADAVOLU, AND J. COOPER (2021): “Using machine learning and qualitative interviews to design a five-question women’s agency index,” Tech. rep., National Bureau of Economic Research.
- JIA, J. AND K. ROHE (2015): “Preconditioning the lasso for sign consistency,” *Electronic Journal of Statistics*, 9, 1150–1172.
- KLINGER, B., A. KHWAJA, AND J. LAMONTE (2013): “Improving credit risk analysis with psychometrics in Peru,” *Inter-American Development Bank. Technical Note No. IDB-TN-587*.
- MCKENZIE, D. (2018): “Can business owners form accurate counterfactuals? Eliciting treatment and control beliefs about their outcomes in the alternative treatment status,” *Journal of Business & Economic Statistics*, 36, 714–722.
- MCKENZIE, D. AND D. SANSONE (2019): “Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria,” *Journal of Development Economics*, 141, 102369.
- MEAGER, R. (2019): “Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments,” *American Economic Journal: Applied Economics*, 11, 57–91.
- (2020): “Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature,” *LSE working paper*.
- MULLAINATHAN, S. AND J. SPIESS (2017): “Machine learning: An applied econometric approach,” *Journal of Economic Perspectives*, 31, 87–106.
- PETERSON, C. (2000): “The future of optimism,” *American psychologist*, 55, 44.
- RIGOL, N. AND B. ROTH (2021): “Strategic disclosure of loan officers and graduation from microfinance: Evidence from Chile,” *Working Paper*.
- WEINSTEIN, N. D. AND W. M. KLEIN (1996): “Unrealistic optimism: Present and future,” *Journal of Social and Clinical Psychology*, 15, 1–8.

Figure 1. Outstanding Debt at ABA Over Time

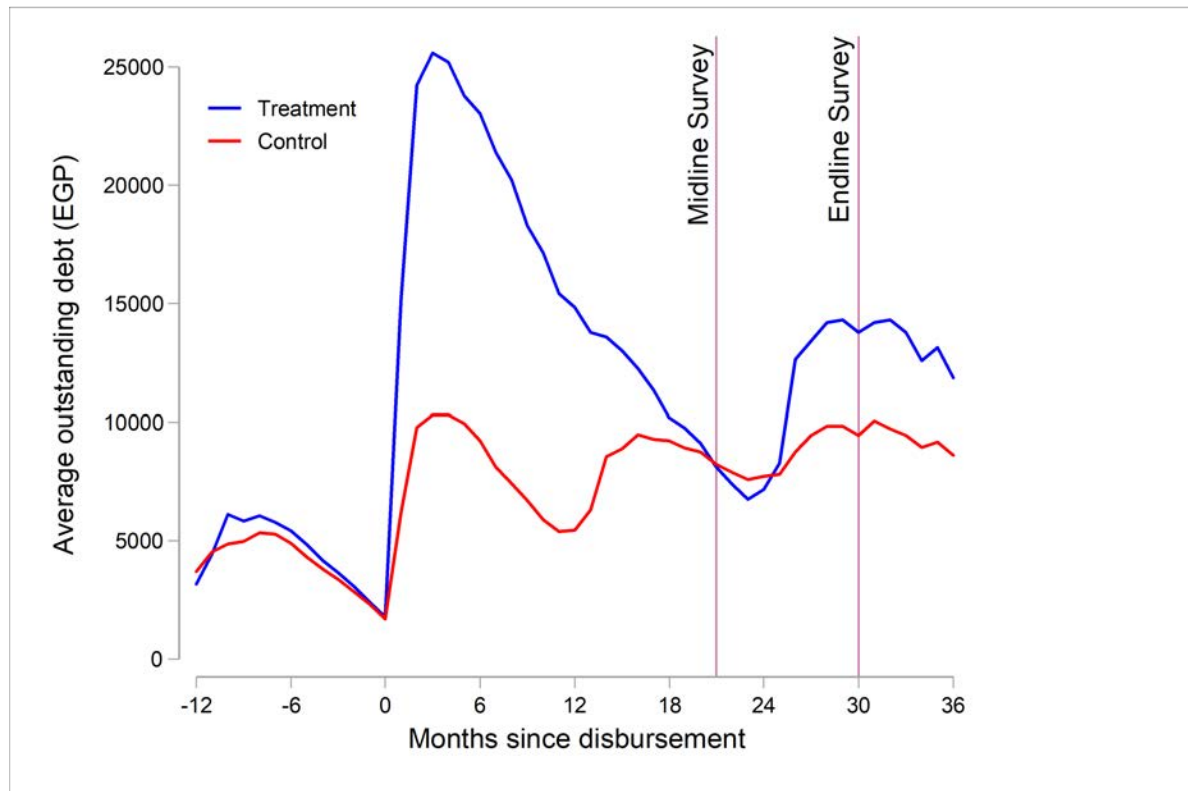


Figure 2. Heterogenous Treatment Effects for Profits

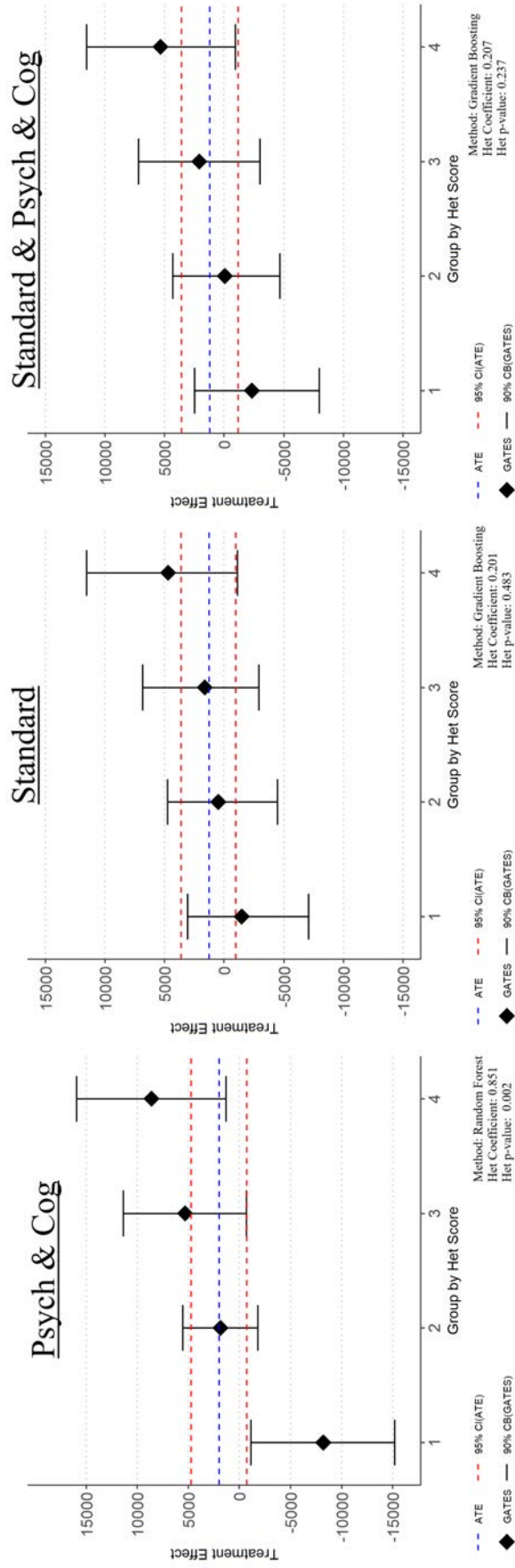


Figure 3. The Model

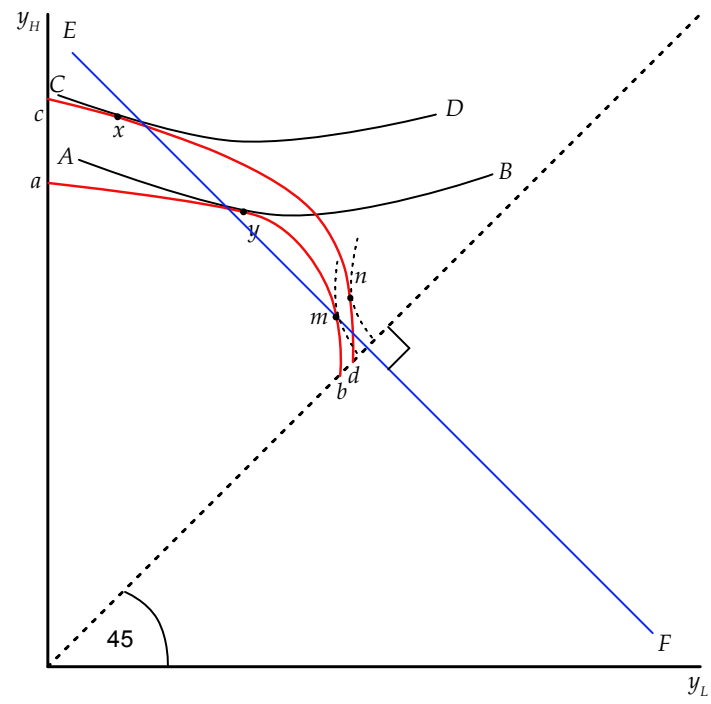


Table 1. Baseline Balance and Take-Up Composition

	Control mean (1)	Control Difference (OLS) (2)	Number of Obs (3)	Predictors of Take Up in Control (4)	Predictors of Take Up in Treatment (5)	p-value for test of (4)=(5) (6)
Baseline Value of Standard Variables						
Age	40.71 {10.0}	-0.49 (0.72)	1003	0.00 (0.02)	0.00 (0.02)	0.81
Female (=1)	0.18 {0.39}	0.03 (0.03)	1004	0.10** (0.04)	0.09*** (0.03)	0.88
Years of education	8.75 {4.52}	0.05 (0.28)	996	-0.02 (0.02)	-0.02 (0.02)	0.90
Business is registered	0.35 {0.47}	-0.01 (0.03)	1004	0.04** (0.02)	0.01 (0.02)	0.24
Years of experience	12.34 {10.1}	0.50 (0.75)	996	-0.03 (0.02)	0.01 (0.01)	0.11
Monthly profits	6810 {9330}	-318 (615)	993	0.05* (0.03)	-0.01 (0.07)	0.47
Monthly revenue	34500 {74011}	-2270 (4260)	993	0.11 (0.14)	0.22 (0.43)	0.80
Monthly expenditures	27800 {69526}	-2110 (4080)	995	-0.12 (0.13)	-0.21 (0.39)	0.83
Monthly wagebill	1410 {3880}	106 (280)	1004	-0.08** (0.04)	0.00 (0.03)	0.09
Has employees (=1)	0.37 {0.48}	0.00 (0.03)	1004	0.07 (0.05)	0.11*** (0.03)	0.47
Number of employees	0.65 {2.19}	-0.05 (0.11)	995	0.03 (0.02)	-0.06* (0.04)	0.03
Size of Previous ABA Loan	7540 {4239}	5.6 (222)	1004	-0.02 (0.03)	0.02 (0.02)	0.15
Have a loan other than ABA (=1)	0.06 {0.24}	-0.03 (0.02)	1004	0.08 (0.12)	-0.04 (0.09)	0.46
Value of other non-ABA loans	2040 {12201}	-805 (936)	1004	-0.04 (0.04)	0.01 (0.01)	0.29
p-value for Joint Test of standard variables		0.787	981			

Notes: Column 1 reports the control group mean with standard deviations in braces. Column 2 reports the coefficient on a treatment indicator from a regression of each row on a treatment indicator and strata fixed effects. Non-binary variables are winsorized at the top 1% level. Columns 4 & 5 report coefficients from a single full interacted regression of a binary for take-up of the loan on each variable listed in the rows interacted with an indicator for control (coefficients in column 4) and an indicator for treatment (coefficients in column 5). For columns 4 & 5 non-binary variables are standardized to simplify interpretation. Standard errors in parentheses. Statistical significance * 0.10; **0.05; ***0.01.

Table 2. First Stage: Effect of Randomized Treatment Status on ABA Borrowing

	Control mean (1)	Treatment - Control Difference (OLS) (2)	Number of Obs
Borrowing Outcomes (First Stage)			
Take-up of Experimental Loan (=1)	0.76 {0.43}	0.09*** (0.03)	1004
Size of Experimental Loan (EGP)	9735 {9423}	10768*** (956)	1004
Loan term (sample: take-up=1), months	13.20 {3.68}	6.49*** (0.34)	810
Size of monthly loan installments	821.5 {939.5}	314*** (58)	1004
ABA outstanding debt ~20 Months after randomization	6673 {11709}	1076 (805)	1004
ABA outstanding debt ~30 Months after randomization	5735 {8375}	1299 (640)	1004
Took another loan after end of experimental loan (=1)	0.48 {0.50}	-0.06** (0.03)	1004

Notes: This table reports the intent to treatment estimates of the impacts of the larger loans. Column 1 reports the control group mean with standard deviations in brackets. Column 2 reports the coefficient on a treatment indicator from a regression that includes strata fixed effects. Standard errors in parentheses. Statistical significance *0.10; **0.05; ***0.01.

Table 3. Effect of Randomized Treatment Status on Repayments to ABA

	Control mean (1)	Treatment - Control Difference (OLS) (2)	Number of Obs
Repayment Behavior			
Eventual Repayment	100% -	0 -	1004
Perfect repayment of each monthly payment	0.76 {0.43}	-0.13*** (0.03)	1004
Total days late over the loan cycle	12.47 {25.68}	14.19*** (2.63)	1004
Total penalty within first 24 months	128 {281}	195*** (48)	1004
Daily value of late payments over the first 24 months	37 {2.6}	29*** (7.6)	1004

Notes: This table reports the intent to treatment estimates of the impacts of the larger loans on repayment behavior. Column 1 reports the control group mean with standard deviations in brackets. Column 2 reports the coefficient on a treatment indicator from a regression that includes strata fixed effects. Standard errors in parentheses. Statistical significance *0.10; **0.05; ***0.01.

Table 4. Primary Impacts on Enterprise Outcomes: ITT and GATES Estimates

	Control Group Mean	Treatment - Control Difference (OLS)	Group Average Treatment Effects (GATES)			
Panel A: Profits	(1)	(2)	(3)	(4)	(5)	(6)
Profits	15,655	1,448 (1,174)	-8,180 ** (3,584)	1,840 (1,872)	5,325 (3,078)	8,611 ** (3,742)
	Control Group Mean	Treatment - Control Difference (OLS)	Conditional Group Average Treatment Effects (CGATES)			
Panel B: Other Primary Outcomes	(1)	(2)	(3)	(4)	(5)	(6)
Revenues	45,626	5,715 (4,415)	-51,463 *** (10,406)	9,453 (7,780)	15,961 (8,035)	47,556 *** (9,481)
Expenses	28,191	5,251 (3,531)	-34,142 *** (8,755)	8,199 (6,352)	11,786 (6,216)	33,940 *** (8,327)
Wage Bill	1,951	144 (247)	-2,041 *** (628)	-561 (481)	874 (545)	2,388 *** (532)
Total Factor Productivity	-0.02	0.04 (0.05)	-0.47 *** (0.11)	-0.17 (0.10)	0.22 (0.10)	0.58 *** (0.10)
Household Expenditure	4,770	446** (211)	-291 (493)	-868 (400)	567 (408)	2,077 *** (470)
	Control Group Mean	Treatment - Control Difference (OLS)	Conditional Group Average Treatment Effects (CGATES)			
Panel C: Secondary Outcomes	(1)	(2)	(3)	(4)	(5)	(6)
Has a Business	0.96	0.00 (0.01)	-0.08 *** (0.02)	0.00 (0.02)	0.03 (0.02)	0.05 (0.02)
Number of Employees	2.55	-0.18 (0.35)	-2.27 *** (0.60)	-1.02 (0.74)	1.27 ** (0.50)	0.84 (0.88)
Value of Assets	361,397	-6,209 (53,336)	-293,152 ** (98,497)	28,058 (93,244)	69,942 (142,199)	128,219 (112,482)
Mental Health Index	0.01	0.00 (0.05)	-0.26 ** (0.10)	0.04 (0.10)	0.17 (0.11)	0.05 (0.10)
Physical Health (out of 5)	2.18	0.04 (0.05)	-0.17 (0.09)	0.00 (0.10)	-0.07 (0.10)	0.11 (0.10)
Total Penalty within 24 months	128	195*** (48)	304 (145)	247 ** (94)	171 * (69)	74 (89)
ABA Loan at 30 Months	0.21	-0.08*** (0.02)	-0.16 *** (0.05)	0.03 (0.05)	-0.22 *** (0.05)	-0.01 (0.06)

Notes: Column 2 reports the ITT estimate from a regression of the outcome variable (row) at follow-up on a treatment indicator. The regression includes controls for strata, survey round (20-month or 30-month), and the baseline value of the outcome when available. Standard errors clustered at the individual level. Columns 3-6 reports results from a two-step process. First, using the random forest algorithm with psychometric, risk, cognitive data, and treatment status, we predict individual treatment effects on business profits. Second, we report the coefficients from a regression of the outcome variable (row) on treatment status interacted with indicator variables associated with which quartile of predicted individual treatment effects (columns) the borrower was assigned to in the first step. For Panel A, this 2nd step is done for profits, the same outcome used to estimate the individual treatment effects. For Panels B and C, this 2nd step is done for other outcomes, but still the quartile assignment is based on predicted individual treatment effects on profits. TFP is the standardized residual from a regression of profits on assets and labor costs. All regressions include controls for predicted profits if in treatment, predicted profits in in control, strata fixed effects, and survey round, and also cluster standard errors at the individual level. Statistical significance *0.10; **0.05; ***0.01 after multiplying all p-values by 4 as part of a Bonferroni correction.

Table 5. Correlates of Enterprise Profits and Total Factor Productivity

Panel A: Business Profits						
Sample Frame: Outcome (EGP/Month):	Control Only		Treatment Only		Full Sample	
	Endline - Baseline Profits		Endline - Baseline Profits		Baseline Profits	
	(1)	(2)	(3)	(4)	(5)	(6)
Standardized Predicted Treatment Effect	-10584*** (1804)	-10607*** (1810)	15893*** (2432)	15866*** (2516)	608 (454)	526 (453)
Years of Education		225 (193)		-185 (298)		205* (117)
Years of Experience		-98 (77)		229 (170)		23 (74)
Number of Observations	906	906	917	917	936	936
Panel B: Total Factor Productivity (Follow-up Surveys Only; Asset Information not Collected at Baseline)						
Sample Frame: Outcome = TFP:	Control Only		Treatment Only		Full Sample	
	(7)	(8)	(9)	(10)		(11)
Standardized Predicted Treatment Effect	-0.284*** (0.044)	-0.283*** (0.044)	0.220*** (0.039)	0.213*** (0.038)	Treatment	0.026 (0.054)
Years of Education		0.001 (0.008)		0.013 (0.008)	Index	-0.284*** (0.044)
Years of Experience		0.009*** (0.004)		0.002 (0.004)	Treatment * ITE	0.499*** (0.058)
Number of Observations	908	908	920	920		1828

Notes: Panel A reports results from regressions of profit outcomes on a standardized measure of the predicted individual treatment effect. Regressions include a control for survey round, and cluster standard errors at the borrower level. Panel B reports results from a similar regression but with total factor productivity in the 20-month and 30-month surveys as an outcome. Statistical significance *0.10; **0.05; ***0.01.

Table 6. Differences in Baseline Characteristics Between Top and Bottom GATES Groups

	Bottom GATES Group	Difference in Top Group		Bottom GATES Group	Difference in Top Group
	Mean {Std. Dev.}	Coeff (s.e.)		Mean {Std. Dev.}	Coeff (s.e.)
Panel A: Loan Officer Perceptions	(1)	(2)	Panel C: Psychometric Data	(3)	(4)
Perceives Large Loan will increase default	0.22 {0.61}	0.24 *** (0.08)	Cognitive & Risk Aversion		
Perceives Large Loan will increase firm revenue	0.24 {0.62}	0.13 (0.08)	Digit Span Recall	2.34 {1.91}	0.88 *** (0.16)
			Ravens Matrices	1.75 {1.54}	0.18 (0.14)
			Financial Literacy Score	5.62 {3.42}	1.49 *** (0.31)
			Hypothetical % allocated to risky investment	0.71 {0.32}	-0.05 * (0.03)
Panel B: Standard Data			Most Predictive Psychometric Measures		
Demographics			I tend to act first and worry about the consequences later	4.32 {1.06}	-0.89 *** (0.09)
Age	41.0 {9.8}	1.72 * (0.89)	I can think of several solutions to any problem	4.67 {0.62}	-0.76 *** (0.06)
Female	0.24 {0.43}	-0.09 ** (0.04)	I prefer to have a flexible schedule- I don't like being tied down	4.45 {0.80}	-0.91 *** (0.08)
Years of Education	8.3 {4.7}	1.06 ** (0.41)	I feel anxious outside my comfort zone	4.11 {1.15}	-0.73 *** (0.10)
Business Characteristics			When I make decisions I usually go with my first, gut feeling.	4.07 {1.25}	-0.23 ** (0.10)
Registered Business	0.36 {0.48}	-0.02 (0.04)	In life, failure is not an option	4.71 {0.56}	-0.70 *** (0.06)
Monthly Revenues	42526 {194315}	-396 (16147)	Optimism Related Psychometric Measures		
Monthly Expenses	49998 {197078}	-745 (16465)	When I make a business decision it is almost always the right decision	4.35 {1.01}	-0.54 *** (0.08)
Monthly Profits	7270 {12200}	-24 (1023)	I have always believed I am going to be successful	4.42 {0.93}	-0.49 *** (0.07)
Has employees (=1)	0.35 {0.48}	0.04 (0.04)	I prefer to focus on opportunities rather than risks	4.34 {0.95}	-0.64 *** (0.09)
Number of Employees	0.59 {1.85}	0.11 (0.16)			
Size of Previous ABA Loan	7725 {4781}	-145 (427)			
Value of Other Loans	5635 {18725}	-3087 ** (1329)			

Notes: Column 1 reports the baseline averages of characteristics for the individuals who are in the bottom group of estimated impacts on profits based on the Chernozhukov (2020) method utilizing psychometric data, while Columns 2 & 4 report the coefficient on an indicator for being in the top group from a regression of the variable in each row and strata fixed effects. Regressions only include people in the top or bottom groups. Panel A includes data from a loan officer survey conducted towards the end of implementation and covers 288 borrowers. Business profits, revenues, expenditures and value of other loans are winsorized at the top 1% level. Hypothetical risk investment question: Respondents asked to allocate 10000 EGP to either a risky (50% chance of doubling/halving) investment or risk-free zero-return investment. Psychometric questions are on a 1-5 scale, with 5=strongly agree & 1=strongly disagree. We report the 6 psychometric variables that were chosen by either a random forest model or a lasso regression of the estimated individual treatment effect on all psychometric variables, as well as 3 additional psychometric questions collected at baseline that we manually identify as being related to optimism. Standard errors in parentheses. Statistical significance *0.10; **0.05; ***0.01.

Appendix Table 1. Loan Size Compared to 7 Other Credit Studies

	Egypt (1)	Bosnia and Herzegovina (2)	Ethiopia (3)	India (4)
Currency	EGP	BAM	ETB	INR
Year of Implementation	2016	2011	2006	2007
Range of loan sizes in local currency	3000-100,000	300-3,000		10,000
Average loan size in treatment group	20,430	1,653	1,200	10,000
PPP conversion factor	2.92	0.71	2.4	11.3
PPP value of the average loan size	\$6,997	\$2,328	\$500	\$885
Annual Value of International Poverty Line (Local Currency)	4,490	1,654	1,500	13,678
Loan size as % of annual poverty line	455%	100%	80%	73%
	Mexico (5)	Mongolia (6)	Morocco (7)	Philippines (8)
Currency	MXN	MNT	MAD	PHP
Year of Implementation	2010	2008	2007	2006
Range of loan sizes in local currency	1,500-27,000	100,000 - 2,000,000	1,000 -39,000	5,000-25,000
Average loan size in treatment group	6,462	502,063	10,571	10,000
PPP conversion factor	8.9	472.5	4.2	16.6
PPP value of the average loan size	\$454	\$1,063	\$2,517	\$602
Annual Value of International Poverty Line (Local Currency)	21,420	794,836	4,514	17,880
Loan size as % of annual poverty line	30%	63%	234%	56%

Notes: Table reports data on the loan size in this study (column 1) and compares it to the loan sizes in seven other randomized experiments on access to credit (columns 2-8). Loan sizes in columns 7 & 8 are medians. PPP conversion rates sourced from the World Bank. Studies sorted alphabetically by country. Set of studies chosen based on inclusion in recent meta-analysis in Meager (2019). Citations, in order by column, are as follows: Augsburg et al (2015); Tarozi et al (2015); Banerjee et al. (2015); Angelucci et al. (2015); Attanasio et al (2015); Crépon et al (2015); Karlan & Zinman (2011).

Appendix Table 2. Balance on Psychometric Variables

	Control mean (1)	Treatment - Control Difference (OLS) (2)	Number of Obs (3)	Predictors of Take Up in Control (4)	Predictors of Take Up in Treatment (5)	p-value for test of (4)=(5) (6)
Baseline Value of Psychometric and Cognitive Variables						
Digit Span Recall	2.79 {1.8}	-0.06 (0.11)	981	-0.03** (0.01)	0.00 (0.01)	0.08
Raven's Matrices Score	1.76 {1.50}	-0.02 (0.10)	981	0.01 (0.01)	0.01 (0.01)	0.99
Hypothetical % allocated to risky investment	0.70 {0.29}	0.01 (0.02)	973	0.04 (0.07)	0.02 (0.05)	0.88
Financial Literacy Score	6.04 {3.4}	0.17 (0.20)	981	0.01* (0.01)	0.01 (0.01)	0.54
In life, failure is not an option	1.68 {0.8}	-0.05 (0.04)	972	-0.01 (0.03)	0.00 (0.02)	0.61
When I make a decision I usually go with my first, gut feeling.	2.03 {1.0}	-0.02 (0.07)	971	0.02 (0.02)	0.00 (0.02)	0.28
I can think of several solutions to any problem	1.75 {0.7}	-0.09* (0.05)	972	-0.01 (0.03)	0.01 (0.03)	0.39
I prefer to have a flexible schedule - I don't like being tied down	1.98 {0.9}	0.01 (0.06)	970	-0.01 (0.03)	0.00 (0.02)	0.48
I tend to act first and worry about the consequences later	2.12 {1.1}	-0.01 (0.07)	972	0.00 (0.02)	-0.01 (0.01)	0.69
I feel anxious outside my comfort zone	2.28 {1.2}	0.01 (0.07)	972	0.00 (0.02)	0.00 (0.01)	0.88
p-value for Joint Test of 50 psychometric & above questions		0.527	978			

Notes: Column 1 reports the control group mean with standard deviations in braces. Column 2 reports the coefficient on a treatment indicator from a regression of each row on a treatment indicator and strata fixed effects. Columns 4 & 5 report coefficients from a single full interacted regression of a binary for take-up of the loan on each variable listed in the rows interacted with an indicator for control (coefficients in column 4) and an indicator for treatment (coefficients in column 5). Standard errors in parentheses. Statistical significance * 0.10; **0.05; ***0.01.

Appendix Table 3. How Survey Attrition Differs by Treatment

	Coefficient (1)	Main Coefficient (2)	Interacted Coefficient (3)		Main Coefficient (4)	Interacted Coefficient (5)
Treatment	0.000 (0.008)	0.007 (0.005)		Age	-0.002 (0.003)	0.006 (0.005)
Digit Span Recall		-0.004 (0.004)	0.008 (0.005)	Female (=1)	0.002 (0.004)	-0.002 (0.006)
Raven's Matrices Score		0.005 (0.003)	-0.001 (0.004)	Years of education	0.006 (0.004)	-0.004 (0.005)
Hypothetical % allocated to risky investment		0.001 (0.005)	-0.002 (0.005)	Years of experience	-0.002 (0.003)	0.002 (0.004)
Financial Literacy Score		-0.001 (0.004)	-0.004 (0.005)	Monthly profits	-0.001 (0.003)	0.001 (0.004)
In life, failure is not an option		-0.004 (0.004)	0.008 (0.006)	Monthly expenditures	-0.013 (0.015)	0.014 (0.018)
When I make a decision I usually go with my first, gut feeling.		-0.003 (0.004)	0.003 (0.006)	Monthly revenue	0.012 (0.016)	-0.012 (0.019)
I can think of several solutions to any problem		-0.001 (0.003)	0.000 (0.003)	Monthly wagebill	-0.002 (0.003)	0.001 (0.003)
I prefer to have a flexible schedule - I don't like being tied down		0.000 (0.004)	-0.001 (0.005)	Business is registered	-0.003 (0.003)	0.001 (0.004)
I tend to act first and worry about the consequences later		0.000 (0.003)	0.001 (0.005)	Has employees (=1)	0.016* (0.008)	-0.018** (0.009)
I feel anxious outside my comfort zone		0.000 (0.004)	-0.001 (0.004)	Number of employees	-0.001 (0.002)	0.005 (0.003)
Size of Previous ABA Loan		0.009 (0.008)	-0.011 (0.007)	Value of other non-ABA loans	-0.010* (0.006)	0.011* (0.006)
				Have a loan other than ABA (=1)	0.029 (0.022)	-0.032 (0.023)
p-value for joint test on treatment interactions for attrition				0.986		

Notes: Column 1 reports the coefficient from a regression of attrition in either survey on the treatment dummy while controlling for survey round and strata (loan officer) fixed effects. Columns 2-5 report the coefficients from one regression. We regress attrition in either survey on a treatment indicator as well as all of the variables in each row (Columns 2 & 4) and an interaction of that variable with treatment (Columns 3 & 5), along with round and strata dummies. All non-binary control variables are standardized to aid in interpretation of the estimates. The last row reports the p-value of a joint test of significant for all the interacted variables in the regression. Standard errors in parentheses. Statistical significance * 0.10; **0.05; ***0.01.

Appendix Table 4. Heterogeneity Using Different Prediction Measures

Heterogeneity Using Loan Officer Predictions				
Panel A: Heterogeneity based on Loan Officer (LO) Perceptions of Repayment				
LO thinks Default will:	Increase with Larger Loan (1)	Stay same with Larger Loan (2)	Decrease with Larger Loan (3)	p-value for (1)=(2)=(3)
Treatment Effects on Profits	880 (2932)	-3127 (3234)	-3365 (13984)	0.65
Treatment Effects on Loan Penalties	20 (59)	48 (62)	314 (225)	0.48
Panel B: Heterogeneity based on Loan Officer (LO) Perceptions of Revenue Changes				
LO thinks Revenue will:	Increase with Larger Loan (1)	Stay same with Larger Loan (2)	Decrease with Larger Loan (3)	p-value for (1)=(2)=(3)
Treatment Effects on Profits	4939 (3232)	-7574* (4140)	3116 (4047)	0.08
Treatment Effects on Loan Penalties	37 (72)	103 (70)	-28 (56)	0.26

Notes: Panels A & B utilize loan officer perceptions collected at baseline on a subset of the sample (N=559 from 288 individuals). Panel A looks at treatment effects split by how the loan officer graded repayment chances in the case of the large loan relative to the small loan. Panel B splits by loan officer grades on how revenues will increase with the large loan relative to the small loan. The regressions also include a survey round dummy and strata fixed effects. Monthly profits in the control group are 8673EGP. Standard errors clustered at the individual level. Statistical significance *0.10; **0.05; ***0.01.

Appendix Table 5. Correlations between 6 Psychometric Variables

	1	2	3	4	5	6
1 I prefer to have a flexible schedule -I don't	1					
2 I tend to act first and worry about the	0.2479	1				
3 I can think of several solutions to any	0.3005	0.1927	1			
4 I feel anxious outside my comfort zone	0.1258	0.183	0.1043	1		
5 In life, failure is not an option	0.2142	0.1668	0.345	0.157	1	
6 When I make a decision I usually go with my	0.1353	0.1677	0.1925	0.1103	0.1608	1

Notes: 16% of respondents answered the same for all of these questions.

Appendix Table 6. What Predicts Performance? Correlates of Individual Treatment Effects

	6 Variable ITE Prediction (1)	Individual Treatment Effect (2)	Individual Treatment Effect (3)
Age	-0.028 (0.037)	-271.6 (198.4)	-205.0 (177.2)
Female (=1)	-0.100 0.083	-433.9 (440.6)	-185.5 (393.7)
Years of Education	0.032 (0.035)	247.5 (186.3)	165.7 (166.9)
Years of Experimence	0.061 (0.036)	133.1 (188.8)	-15.1 (169.3)
Monthly Profits	0.032 (0.052)	-186.9 (274.4)	-264.1 (245.2)
Business is Registered	-0.015 (0.034)	35.0 (177.6)	64.3 (158.8)
Has employees (=1)	-0.032 (0.037)	182.4 (196.9)	262.0 (175.9)
Value of other non-ABA loans	-0.037 (0.031)	-376.8* (161.5)	-287.4 (148.2)
Size of Previous ABA Loan	0.016 (0.037)	560.564** (195.927)	532.1** (175.7)
Monthly Expenses	0.236 (0.242)	1122.0 (1281.9)	548.1 (1144.5)
Monthly Revenue	-0.229 (0.258)	-1075.1 (1367.3)	-520.4 (1220.6)
Monthly Wagebill	-0.059 (0.040)	-395.7 (211.2)	-248.9 (190.5)
6 Variable ITE Prediction			0.984*** (0.062)
Number of Observations	976	936	936
p-value on joint test	0.687	0.000	0.000

Notes: Outcomes and non-binary dependent variables standardized. 6 variable ITE prediction is generated by regressing the ITE on the 6 variables outlined in Panel C of Table 6, and then predicting the value for each borrower using the estimated regression coefficients. Statistical significance *0.10; **0.05; ***0.01.

Online Appendix

Machine Learning Appendix

This appendix details the machine learning procedures for estimating treatment effect heterogeneity. We mostly follow the generic machine learning techniques of [Chernozhukov et al. \(2020\)](#), with a couple modifications explained below.

To build some intuition for the method, suppose that, based on prior knowledge or theory, we were able to predict a treatment and a control outcome for each of our firms, based only on baseline data. We would then be able to subtract these two outcomes to form a predicted individual treatment effect (ITE) for each participant in our study. We could then use our experimental data to test whether the predicted ITE accurately captures heterogeneity in experimental treatment effects. For example, we might run a linear regression including the predicted ITE, interacted with treatment, and test whether the resulting coefficient is equal to zero. If the coefficient was greater than zero, this would imply that the predicted ITE captures some of the actual treatment effect heterogeneity, and if it was equal to 1 then it would perfectly capture it. Because the experimental data has not been used to generate the predicted ITE, there are no concerns about overfitting, and standard hypothesis testing methods can be used.

In practice we do not have sufficient prior knowledge to undertake this task, but we do know that, with sufficiently rich baseline data, machine learning methods are good at producing well performing predictive models. This suggests using a subset of our experimental data (a training set) to train a predictive model, and then testing for heterogeneity using the remaining part of the data (a testing set). Again, because the testing set has not been used to train the model, overfitting is not a concern. Of course, the results of this method would depend on the exact split of the data used, which suggests trying multiple splits and then taking steps to correct for the number of splits used. We would also like to account for the fact that the researcher may have used multiple different machine learning approaches and chosen the one with the “best” performance.

An important implication of this discussion is that the specific machine learning algorithm used is not relevant, it only matters that the model is successful at predicting outcomes using baseline data. This is the logic, that we follow, put forward in [Chernozhukov et al. \(2020\)](#). We use four different and common machine learning algorithms, but we could use any algorithm with good predictive performance. A second implication is that, while it may not be possible to do correct inference for the entire conditional average treatment effect (CATE) function, it may be relatively straight forward to do correct inference for specific features of the CATE, as suggested by the linear heterogeneous treatment effect example above. A major contribution of [Chernozhukov et al. \(2020\)](#) is to provide a procedure that accounts for multiple sample splits and multiple machine learning methods and provides correct inference around parameters or features that matter for the analysis of heterogeneous treatment effects. In our case we care about two primary parameters, (1) whether the data show evidence of important heterogeneity in treatment effects (BLP), and (2) how the treatment effect differs for different groups in the sample (GATES).

For the first parameter the procedure calls for estimating a model for the individual treatment effect using a random half of the data (training set), and then regressing that value interacted with treatment on the actual outcomes in the other half of the data (testing set). If the models are predictive then when we run regression 1 we would find that the coefficient on the predicted individual treatment effect, interacted with treatment, will be positive and statistically significant. This would mean that individual treatment effect estimated for the people in the testing test (using the model generated from the data for the people in the training set) is indeed predictive of the actual outcomes from the experiment.

Part of the inference procedure includes rerunning the analysis 100 times with different splits of the data into training & testing sets and taking the median values of the coefficients to protect against the uncertainty of the data splitting. To account for this splitting and rerunning confidence intervals are considered to be at the $1-2 * \alpha$ instead of $1-\alpha$ level. To account for the use of multiple machine learning models, a Bonferroni

adjustment is used.

For the second parameter, the predicted ITE is sorted for everyone in the testing set, and then they are split into quartiles. This is followed by a fully interacted regression of treatment and the set of group estimated quartile bins, so that we can estimate what the treatment effect is for each group. Again, confidence intervals are at the $1-2 * \alpha$ level to account for sample splitting and because the method initially generates predictive models using four different machine learning methods it multiplies all p-values by four, i.e. a Bonferroni correction to conservatively control for multiple hypothesis testing. We report results from the method that performs best (i.e. has the highest coefficient from regression 1), and with this Bonferroni adjustment we should not need to worry about false positives. Nonetheless we also find heterogeneity using other methods. But looking for heterogeneity using less predictive algorithms is like estimating a causal effect using an instrument with a less powerful first stage - you could do it but a researcher would naturally prefer the stronger option, for both statistical and theoretical reasons.

[Chernozhukov et al. \(2020\)](#) estimates the GATES parameters using only the testing half of the sample. In each of the 100 splits of the data, they generate ITEs for half of the sample, and then run their regressions on that half of the sample. They do this 100 times and then take the median coefficient and median standard error from the 100 regressions. We follow this procedure for our GATES estimates on profits (panel A in Table 4). In our CGATES estimates (panels B & C in Table 4) we deviate slightly by taking the median ITE from all 100 splits of the data for each person and then allocate them to groups based on this value. We then run the fully interacted regression of these groups and treatment to recover the group average treatment effects. This allows us to utilize the full sample, while ensuring that the estimates remain “honest”, since no person’s data went into estimating their own ITE.

Research degrees of freedom: While the method is relatively mechanical, researchers have some degree of freedom which could allow them to cherry-pick results. Researchers can choose which data to include in the analysis, which machine learning

algorithms to try out, what proportion of the data to put into training vs testing sets, and the number of groups in the GATES procedure. In our setting we show how we change the baseline data by including only psychometric & cognitive data in one case, standard data in another, and all data in a third. We use only the four machine learning algorithms that are chosen by [Chernozhukov et al. \(2020\)](#), and we correct for the multiple methods as noted above. We split testing and training data in half just like the initial authors, and believe that an alternative approach would seem manufactured. We do diverge by using four groups instead of five in the GATES analysis. This does not change the estimate of whether heterogeneity exists (BLP), but provides us more power when considering impacts on each group. Estimates are economically similar, but less precise when we use five groups.

Who are in the different groups? A limitation to these methods is that the models produced by the algorithms are not easy interpreted. While the models can do a good job predicting the outcomes for people outside the sample, the model itself is not informative about why some people are predicted to do well and others are predicted to do badly. To address this [Chernozhukov et al. \(2020\)](#) suggests using “CLAN Analysis” which compares the baseline characteristics of people in the top group and those in the bottom group. But it’s not clear which characteristics are worth comparing, which is a difficult problem when there are many baseline covariates that the algorithms could be combining non-parametrically.

To address this concern we take the median value of the estimated individual treatment effect across the 100 splits of the data, and then attempt to see which baseline variables are most predictive of the estimated effect. We do this in two ways, once using a lasso, and another time using a random forest algorithm. For the lasso we utilize cross validation to estimate the appropriate lambda which yields a model with five variables. The choice of five is in line with other recent work utilizing machine learning techniques to generate summary measures ([Jayachandran et al., 2021](#)).

Ethics Appendix

We have developed this appendix in an effort to describe the ethical considerations of this experiment, and clarify the nature of the collaboration between the researchers and ABA. We follow the framework put forth in [Asiedu et al. \(2021\)](#), for the sake of comparability within economics.

1. **Equipoise** As we describe in the paper, there was significant hesitancy from bank management in providing these larger loans to clients. We used the evaluation to provide the lender an opportunity to test whether or not these loans would be beneficial to their clients as well as to the bank's bottom line.
2. **Role of Researchers with Respect to Implementation:** Bryan, Karlan and Osman are not active researchers in the project. Bryan, Karlan and Osman designed the randomization and evaluation protocols and managed the data collection activities and all of the data analysis. ABA, and not the researchers, designed and implemented the treatments (i.e., the loans).
3. **Potential Harms to Research Participants from the Interventions:** We designed the study to limit the potential harms to participants. In particular participants voluntarily applied for a larger loan, and were approved by their loan officer and the bank's credit committee. All participants expected to benefit from the intervention, and all the bank staff also expected participants to benefit. Our results show that some did not benefit, leading to decreases in business profits and related outcomes.
4. **Potential Harms to Research Participants from Data Collection or Research Protocols** We do not believe participants were subjected to any harms from data collection. Participants were able to refuse to answer any questions they wanted to and were told they could stop the interview at any time.
5. **Financial and Reputational Conflicts of Interest:** Bryan, Karlan and Osman did not receive any form of financial compensation as part of this study (nor did any

assistants or staff associated with the research team). No employee of ABA was named as a PI or participant in any research grant that provided funding for this project.

The research questions pursued in this study and the results described in this study are novel and different form of prior work conducted by the authors. We perceive no reputational conflicts of interest.

6. **Intellectual Freedom:** This study was conducted through a collaboration between PIs and the Alexandria Business Association. The study was conceived and designed by the PIs, who maintained full intellectual freedom throughout all stages of the project. At no point did the partner have undue influence on the analysis or interpretation of results.
7. **Feedback to Participants or Communities:** We intend to share our results with participants via email after our work is subject to peer-review.
8. **Foreseeable Misuse of Research Results:** The authors recognize that the results are relevant for public policy and regulatory activities in credit markets. While misinterpretation or deliberate mischaracterization of the results could have implications for individuals, communities and firms, we are unaware of any set of outcomes or analyses that ex-ante would be predicted to favor advantaged parties. In fact, our search for heterogeneity is partly motivated by this kind of concern regarding studying merely the “average” impact of credit for firms, since average results may mask the presence of both winners and losers.

Survey Questions

Risk Questions

R1: How do you see yourself? Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please select one number on the scale, where 0 equals “unwilling to take risks” and 10 equals: “fully prepared to take risks”

R2: Imagine you have EGP 10,000 to invest. You may choose to “invest” all or a portion of your money. There is a 50% chance to double the amount of money you choose to “invest”. It is equally possible that you could lose half the investment. For example, if you choose to invest EGP 1,000, there is a 50-50 chance of getting EGP 2,000 or EGP 500. How much of the EGP 10,000 do you invest?

Financial Literacy Questions

FL1: As the demand for a good or service rises, all other things being equal, its price is likely to:

Fall; Rise; Stay the same; There is not enough information to know; I don't know;
Refused to answer

FL2: The amount by which revenue from sales exceeds costs in a business is known as:

The profit margin; The cost of capital ; Owner's equity; Gross sales; I don't know;
Refused to answer

FL3: As the price of raw materials used to produce a product fall, all other things being equal, the profit margin on sales of that product will:

Fall; Rise; Stay the same; Remain unchanged ; I don't know; Refused to answer

FL4: Monies that go in and out of business over a period of time are known as:

Cash flow; Sales; Assets; Liabilities ; I don't know; Refused to answer

FL5: A sum of money that is owed or due is known as:

A credit; Equity; A debt; A grant; I don't know; Refused to answer

FL6: Annual sales net of all discounts and sales taxes is known as:

Profit; Turnover; Gross sales ; Cash flow; I don't know; Refused to answer

FL7: A sum of money paid regularly (typically annually) by a company to its shareholders out of its profits (or reserves) is known as:

Interest; Profit; A loan; A dividend; I don't know; Refused to answer

FL8: You export the majority of the goods you sell. All other things being equal, if the value of your own currency falls (depreciates) relative to the currency of the market you export to, the goods you sell to that market will be:

More expensive; Cheaper; The same price as before; There is not enough information to know; I don't know; Refused to answer

FL9: You import raw materials to produce the goods you sell. All other things being equal, if the value of your own currency rises (appreciates) relative to the currency you import raw materials from, the cost of producing the goods you sell will:

Rise; Fall; Stay the same; There is not enough information to know; I don't know; Refused to answer

FL10: Debts or financial obligations incurred during business operations are known as:

Equity; Credit; Assets; Liabilities ; I don't know; Refused to answer

FL11: A financial report summarizing a company's assets (what it owns), liabilities (what it owes) and owner's equity at a given time is known as:

The Bought Ledger; The Balance Sheet; The Cash flow Statement; The Profit & Loss Statement; I don't know; Refused to answer

FL12: A financial statement summarizing a company's performance and financial po-

sition by reviewing revenues, costs and expenses during a specific period of time is known as:

The Bought Ledger; The Balance Sheet; The Cash flow Statement; The Profit & Loss Statement; I don't know; Refused to answer

FL13: The total assets of a company, minus its total liabilities represent the capital available for distribution to its shareholders. This is known as:

Profit; Owner's equity; Loss; Debt; I don't know; Refused to answer

FL14: The risk of a customer not paying for goods they have received is an example of:

Financial risk; Operational risk; Strategic risk; Compliance risk; I don't know; Refused to answer

FL15: Which of the following is not included in a cash flow control statement?

Cash coming in; Assets; Cash going out; Cash balance; I don't know; Refused to answer

FL16: Which of the following is not an operating expense?

Payroll; Taxes; Supplies; Dividend payments ; I don't know; Refused to answer

Psychometric Questions

Baseline

At work, I need to be in control

I always say what I think

I don't follow, I lead

I will do anything to get what I want

When I make a business decision it is almost always the right decision

I have always believed I am going to be successful

Deep inside, I know I am better than most people

I don't get the recognition I deserve
People often tell me how great I am at what I do
I've got a great sense of humour
It is always better to be in the background than in the centre of attention
Modesty gets you nowhere
When I make decisions I usually go with my first, gut feeling
I prefer to focus on opportunities rather than risks
Without risk there is no reward
I tend to act first and worry about the consequences later
It's always good to question authority
I find it difficult to take orders from other people
I only trust myself
People often struggle to understand my ideas
I see patterns and connections where other's don't
I always know when to give up, and move on to something else
I would work seven days a week if I could
I can't wait to get out of bed in the morning - there is always so much to do
In life, failure is not an option
I am critical of myself
I feel anxious outside my comfort zone
Success is never down to luck
Whenever I cross something off my "to-do list", I add something new straight away
You should never take shortcuts in life
I stay calm even during a crisis
I can think of several solutions to any problem
I prefer to have a flexible schedule - I don't like being tied down
I always get things done ahead of time
When I need to, I act quickly without thinking too much
I plan everything

I am more concerned about getting the job done than following office rules
I can concentrate well even when my office is messy
I always check and double check my work
I can leave work unfinished and move on
I am more concerned about the big picture than the details
I keep my promises
I see business opportunities where others don't
It is not that I don't see profitable business opportunities, I just don't have the
motivation to do anything about them
I spend a lot of time planning for my future
I have a strong desire to be successful in life
I am a results oriented person
I am a very competitive person
Some people think I am lazy
I am always trying to improve my performance, whatever I am doing

Follow Up 1

Discussion is fine, as long as I make the final decision
I don't like being told what I can and cannot do
People think twice before confronting me
In years time, I want people to still be talking about the great work I have done
As a child I always kept my parents on their toes
I am a very easy person to get along with
I am a dreamer
I often have too many ideas in my head at once
I get annoyed when people do not take their work seriously
I like it when people are straight with me and give honest feedback
I feel in complete control of my life
I hate letting people down

I have practical skills that others don't
People often ask for my opinion on business matters
I know that others may be better at a task than I am
I create my own path in life
I choose my words carefully
I like being spontaneous
I make decisions quickly and move on
I do my best work when I leave things to the last minute
If I disagree with someone, I tell them what I think
I am good at making last minute changes to plans
When working in teams, I tend to come up with more ideas than others
I am always trying to find new ways of doing things
People who believe in a better future are just naive
Great business ideas change the world
When it comes to exploiting business opportunities, I am often too cautious
Even if I know how to do something, I would always try to do it in a different way
Every decision I make is part of my long-term career plan
I don't have any big plans for the future
I need to do my work exactly right
I get annoyed when plans fall through
I have a brain for business
I always know when there is a gap in the market for a new product
In my group of friends I am the most creative person
I am always willing to take financial risks
When investing my money, I would rather be safe than sorry
I am someone who likes to win at whatever I do

Follow Up 2

Status is the most important thing in life

I would love to be famous
People are jealous of me
I am naturally impatient
I rarely recognize valuable business openings unless they are really obvious
I am a natural risk taker
I was in trouble a lot at school
I was in trouble a lot when I was a kid
I do things my way
I often hesitate to act on valuable business opportunities
I don't like following rules
I sometimes feel overwhelmed by my own creativity
I am an 'ideas person' details bore me
Being different is necessary for success
Before I go to sleep I think, "what could I have done better today?"
I like to live dangerously
As a child, I was never interested in school
It's my way or the high way
I do not let other people's opinions affect me
I am excellent at what I do
I handle difficult tasks well
There is no challenge I feel I can't overcome
I can easily deal with unexpected events
I am more optimistic than most people I know
I prefer to do things in ways that are well established
My judgements can be wrong
I always stick to the rules
I am so busy I sometimes forget things
I need everything to be just right
I change and update my plans constantly

I act exactly according to plan
I will risk making mistakes to get things done
I am always looking out for new business opportunities
I think my decisions through carefully
I feel great after organising my workspace
I always follow directions
I see commercial opportunities in everything
I don't like rushing into things
Some people think I am risk averse
I like to be prepared at all times
I am quick to spot ways of making money
Other people think I am highly creative
If there is a profitable business opportunity, I rarely miss it
I prefer to solve problems in novel ways
It is always best to follow rules and social norms
I am always trying to make things better
I find it easy to apply my creativity in everyday life
I rarely see good business opportunities, even in my area of expertise
My aim in life is to find new ways to make money
Even when I spot a profitable business opportunity, I rarely act on it
I am not afraid to take business risks
I often solve problems in unusual ways

Cognitive Questions

Digit Span Recall

Now we're going to play a game. A series of digits will appear for a few seconds. Try to remember the digits and repeat them back in order. After each successful turn an extra digit will be added. [Max at 10].

Raven's Matrices

Fluid Intelligence: Raven's Matrices
240 Seconds to complete

Item 1

Answer options

a	b
c	d
e	f
g	h

Item 2

Answer options

a	b
c	d
e	f
g	h

Item 3

Answer options

a	b
c	d
e	f
g	h

Item 4

Answer options

a	b
c	d
e	f
g	h

Item 5

Answer options

a	b
c	d
e	f
g	h

Item 6

Answer options

a	b
c	d
e	f
g	h

Item 7

Answer options

a	b
c	d
e	f
g	h

Item 8

Answer options

a	b
c	d
e	f
g	h

Item 9

Answer options

a	b
c	d
e	f
g	h

Item 10

Answer options

a	b
c	d
e	f
g	h